



**Universidad
Politécnica
de Cartagena**



industriales
etsii UPCT

Estudio de Dependencias en los Mercados Eléctricos

Titulación: Ingeniería en Organización Industrial

Alumno/a: Eufemio Pacheco Rocamora

Director/a/s: Dra. María del Carmen Ruiz Abellón

Dr. Antonio Gabaldón Marín

Cartagena, 27 de Septiembre de 2012





RESUMEN

Desde comienzos de los años 90, viene produciéndose el fenómeno de liberalización del sector eléctrico en numerosos países. Por ejemplo, la liberalización del sector eléctrico español comenzó el 1 de enero de 1998 mediante la entrada en vigor de la Ley del Sector Eléctrico 54/1997, de 27 de diciembre.

Entre las consecuencias más importantes que supone el fenómeno de liberalización, podemos destacar la introducción de la competencia en las actividades de generación y comercialización, la aparición de un mercado de electricidad regido por principios de competencia como referencia fundamental para establecer el precio de electricidad (*pool* eléctrico) y, por último, la progresiva liberalización del consumo, permitiendo a los consumidores finales tener libertad de elección para el abastecimiento de sus necesidades de energía eléctrica.

Teniendo en cuenta que el proceso de liberalización es un fenómeno relativamente novedoso, existe un gran interés tanto desde el punto de vista académico como desde el propio mercado en el análisis de las series de precios de la electricidad, no disponiendo de demasiada literatura previa al respecto.

Concretamente, en este Proyecto nos planteamos estudiar las distintas dependencias o asociaciones que existen dentro y entre los distintos mercados eléctricos. Por ejemplo, analizamos el grado de asociación que tiene la serie de precios de un mercado en el instante actual con los precios casados en instantes pasados. Además, estudiamos si la evolución de dicha asociación es la misma en los mercados eléctricos analizados o si por el contrario se advierten comportamientos distintos dependiendo de la época del año. Por otra parte, analizamos el grado de asociación que presentan los precios entre los distintos mercados eléctricos, europeos y no europeos, destacando cuándo se consideran asociaciones estadísticamente significativas.

En el capítulo uno, se introducen las nociones básicas para la comprensión del funcionamiento de los mercados eléctricos, así como las consecuencias de la liberalización de los mercados eléctricos. Además se describe brevemente cómo se produjo ese proceso en distintos países.

En el segundo capítulo, distinguimos varias secciones. La primera sección está dedicada a introducir el proceso de codificación de una serie y definir el concepto de entropía de permutación, así como algunas de sus aplicaciones. En la segunda sección, se describe la utilidad de las tablas de contingencia para detectar asociación o dependencia entre variables y se introducen algunas medidas de asociación entre variables. La sección tres está dedicada a introducir nuestra propuesta de medidas de asociación y de “distancia” entre dos series temporales, mediante el uso de las permutaciones.

El capítulo tres está dedicado a la implementación de la metodología propuesta usando MatLab, mostrándose además los manuales de usuario de cada uno de los programas desarrollados.



En el capítulo cuarto, estudiamos la eficiencia del método propuesto usando datos simulados. En la primera parte, se proponen pares de series, algunos dependientes y otros independientes, para analizar la fiabilidad de nuestras medidas en la detección de asociaciones. En la segunda parte, se analiza y discute cómo afecta la presencia de estacionalidad en el estudio de la dependencia de series. En la última parte del capítulo se muestra cómo podemos identificar el retardo en el tiempo que proporciona mayor dependencia funcional entre las series.

Por último, en el capítulo cinco, se incluyen los resultados de aplicar la metodología propuesta (cálculo de medidas de asociación y de distancia por ventanas móviles) a series de precios de distintos mercados eléctricos. Se distinguen dos secciones. La primera, está dedicada al estudio de autodependencias en los mercados eléctricos, mientras que la segunda está dedicada al estudio de dependencias entre distintos mercados eléctricos. En esta parte se incluye cómo quedaría la clasificación de los mercados analizados en clusters homogéneos, usando para ello métodos jerárquicos de clasificación con distintos tipos de enlace y varias distancias.



ÍNDICE GENERAL

	<u>Nº Pag.</u>
1. La liberalización de los mercados eléctricos	1
1.1. Introducción	1
1.2. Funcionamiento del mercado	2
1.2.1. Mercado Spot	2
1.2.2. Mercados de Ajustes o “Intradiarios”	3
1.2.3. Mercados de Futuros	3
1.2.4. Mercado a Plazos (Contratos Forward)	4
1.2.5. Contratos de Opciones	5
1.2.6. Operador de Sistema	5
1.3. Algunos mercados eléctricos europeos	6
1.3.1. España	6
1.3.2. Austria	6
1.3.3. Italia	7
1.3.4. Francia	7
1.3.5. Nordpool	8
1.3.6. Alemania	9
1.4. Mercados eléctrico internacionales	11
1.4.1. Ontario	11
1.4.2. PJM	12
1.4.3. Australia	13
2. Cómo detectar dependencias entre series usando permutaciones	16
2.1. La entropía de permutación	17
2.2. Las tablas de contingencia y medidas de asociación entre variables	22
2.3. Medidas de “asociación” y “distancias” entre series usando permutaciones	38
3. Software utilizado	44
3.1. Manual de usuario del programa Dependencias	44
3.2. Manual de usuario del programa Retardo	49
3.3. Manual de usuario del programa Mostrar	50
3.4. Implementación de la metodología en Matlab	50
3.4.1. Programa Dependencias	50
3.4.2. Programa Retardo	54
4. Funcionamiento de las medidas propuestas para datos simulados	55
4.1. Eficacia en la detección de dependencia funcionales	55
4.2. Estudio de la dependencia ante la presencia de estacionalidad	67
4.3. Identificación del retardo que proporciona máxima dependencia	70
5. Aplicación a los mercados eléctricos	80
5.1. Autodependencias en mercados eléctricos	80
5.2. Dependencias entre precios de distintos mercados	94
6. Conclusiones y líneas futuras	115



7. Bibliografía	117
I. Anexo	119
a. Programación Matlab	119
i. Función apilar	119
ii. Función desestacionalizar	119
iii. Función ordenar	120
iv. Función prim	120
v. Función tablabidim2	121
vi. Función tablabidretardo	126
b. Matrices de distancia	128
c. Dendogramas	137



1. LA LIBERALIZACIÓN DE LOS MERCADOS ELÉCTRICOS

1.1. INTRODUCCIÓN

Durante la década de los noventa se inició el proceso de liberalización de la mayoría de los mercados eléctricos europeos. Concretamente en el año 1996, siguiendo el ejemplo de Estados Unidos y del Reino Unido, la Unión Europea adoptó las leyes que obligaban a los estados miembros a comenzar un proceso paulatino de liberalización de sus mercados eléctricos, el cual se inició, en la mayoría de casos, a partir del año 1998. Anteriormente, pocas o incluso una única empresa controlaba todos los aspectos del mercado eléctrico, desde la generación hasta la distribución. Con esta ola de cambios y nuevas regulaciones, los mercados han pasado a estar formados, actualmente, por miembros tan diversos como empresas generadoras, empresas distribuidoras o consumidoras y el operador del mercado. Las empresas generadoras ofrecen periódicamente una cierta cantidad de producción a un precio según la estrategia de la compañía. Por otra parte, las empresas distribuidoras o consumidoras realizan ofertas de compra según una cierta previsión de su demanda.

En enero 1998 se inició este proceso en España, tras la aprobación de la Ley 54/1997, de 27 de noviembre. La presente Ley tiene como fin básico establecer la regulación del sector eléctrico, con el triple y tradicional objetivo de garantizar el suministro eléctrico, garantizar la calidad de dicho suministro y garantizar que se realice al menor coste posible, todo ello sin olvidar la protección del medioambiente.

El propósito liberalizador de esta Ley no se limita a acotar de forma más estricta la actuación del Estado en el sector eléctrico. En la generación de energía eléctrica, se reconoce el derecho a la libre instalación y se organiza su funcionamiento bajo el principio de libre competencia. La retribución económica de la actividad se asienta en la organización de un mercado mayorista. Se abandona el principio de retribución a través de unos costes de inversión fijados administrativamente a través de un proceso de estandarización de las diferentes tecnologías de generación eléctrica.

El transporte y la distribución se liberalizan a través de la generalización del acceso de terceros a las redes. La propiedad de las redes no garantiza su uso exclusivo. La eficiencia económica que se deriva de la existencia de una única red, raíz básica del denominado monopolio natural, es puesta a disposición de los diferentes sujetos del sistema eléctrico y de los consumidores. La retribución del transporte y la distribución continuará siendo fijada administrativamente, evitándose así el posible abuso de las posiciones de dominio determinadas por la existencia de una única red. Asimismo, para garantizar la transparencia de esta retribución, se establece para las empresas eléctricas la separación jurídica entre actividades reguladas y no reguladas en cuanto a su retribución económica.



Con estas medidas se conseguía mantener reguladas las actividades de transporte y distribución, dada su característica de monopolios y a la vez, se liberalizan y diferenciaban las actividades de generación y comercialización. De este modo, se evitaba que ninguna sociedad pudiera ejercer a la vez dos actividades diferentes; es lo que se conoce como desintegración vertical.

1.2. FUNCIONAMIENTO DEL MERCADO

La introducción de competencia en el sector hace necesarias nuevas instituciones que no existían en el modelo tradicional. Por un lado, es común la aparición de uno o varios Operadores de Mercado, encargados de facilitar la compraventa de energía entre los agentes, gestionando un mercado organizado u ofreciendo productos de derivados eléctricos. Por otro, la creación del Operador del Sistema, que permite operar el sistema de forma centralizada en condiciones compatibles con los resultados del mercado. Ambas instituciones representan, respectivamente, la coordinación económica y técnica de los recursos del sistema. Además, se hace necesaria la existencia de un regulador independiente encargado de resolver los conflictos que surjan entre los agentes.

En el marco tradicional, la programación horaria, o Unit Commitment, utilizaba modelos de optimización para decidir en un horizonte diario o semanal los generadores que debían producir y la cantidad de energía a producir para satisfacer la demanda de la forma más eficiente. La liberalización del sector ha transformado este proceso en una subasta, o varias con varios horizontes temporales, realizada por el OM.

1.2.1. Mercado Spot

La función característica del OM es la gestión del mercado spot o mercado organizado de corto plazo, en el que se suele realizar una subasta de la energía a generar o consumir con un horizonte temporal de, generalmente, un día. El OM determina las ofertas aceptadas tanto de compra como de venta de energía, así como el precio del sistema en cada periodo (normalmente horario o semihorario) del día siguiente. El modelo de casación puede ser cualquiera de los vistos en el apartado anterior.

El mercado spot es el núcleo central del modelo liberalizado y su diseño es clave en el éxito de un mercado eléctrico competitivo. Un deficiente diseño unido a una regulación incorrecta puede llevar al fracaso de todo el proceso de reformas.

Los primeros mercados eléctricos organizados tenían las siguientes características principales:

- Papel pasivo de la demanda: los consumidores son simples tomadores de precio.
- Obligación de acudir al mercado: los generadores sólo pueden negociar su energía a través del mercado organizado.



- Modelos de casación compleja: utilización de algoritmos de optimización en la casación de oferta y demanda similares a los algoritmos del marco regulatorio tradicional pero adaptados para tener en cuenta las ofertas de los agentes.

Los modelos más recientes, sin embargo, presentan características diferentes:

Papel activo de la demanda: la demanda puede participar en el mercado en condiciones de competencia, al igual que la generación.

Participación voluntaria en el mercado: los agentes pueden negociar libremente su energía a través del mercado organizado o, alternativamente, establecer contratos bilaterales físicos en las condiciones que libremente elijan.

Modelos de casación y formatos de presentación de ofertas más simples, obligando a los agentes a internalizar las restricciones técnico-económicas en sus ofertas.

Una posible tendencia futura es el nuevo concepto de mercado establecido en, el NETA. El NETA lleva al extremo la posibilidad de los agentes de establecer contratos bilaterales, suprime el mercado spot y todos los agentes que desean comprar o vender electricidad lo hacen libremente estableciendo contratos. Se considera que esta negociación puede llevarse a cabo bien mediante mercados organizados (existen varios Operadores de Mercado) o mediante contratos bilaterales.

1.2.2. Mercados de Ajustes o “Intradiarios”

Son llamados de mercados de ajustes o “intradiarios” porque permiten a los agentes corregir posibles errores de internalización cometidos en los mercados de más largo plazo, o correcciones de la previsión de generación/consumo mediante transacciones en estos mercados de más corto plazo (inferior al día cuando el horizonte temporal del mercado spot es diario). El Operador del Mercado es normalmente el encargado de gestionar estos mercados.

1.2.3. Mercados de Futuros

Un contrato de futuros es un acuerdo para comprar o vender energía eléctrica a un cierto precio en una fecha futura dada. Del agente que realiza la compra del contrato de futuros se dice que se encuentra en una “posición larga” mientras que del agente que realiza una venta de futuros se dice que se encuentra en una “posición corta”.

Una primera clasificación de los mismos puede establecerse atendiendo al hecho de que dichos contratos obliguen a la entrega física de dicha energía:

- Contratos con obligación de suministro físico, en cuyo caso el agente debe producir o consumir físicamente la energía a que hace referencia el contrato en las condiciones y periodos estipulados.
- Contratos sin obligación de suministro físico (contratos financieros). Los contratos de futuros financieros que se inician no concluyen con la entrega de la energía eléctrica porque los agentes cierran sus posiciones antes del periodo de



entrega especificado en el contrato. La liquidación de una posición supone realizar una transacción contraria a la original. Por ejemplo, si un agente compra (vende) en mayo 20 contratos de futuros mensuales en carga base (un contrato para todas las horas del mes) para el suministro de dicha energía en el mes de julio, puede liquidar su posición en junio, si no tiene interés en realizar la transacción de dicho producto en el periodo estipulado, vendiendo (comprando), es decir tomando posición corta (larga), 20 contratos iguales. La diferencia está en que el precio del producto comprado en mayo no tenía la misma valoración en junio, debiendo por lo tanto pagarse la diferencia económica de valoración entre el momento de compra y el de venta.

La calidad de suministro suele estar sometida a la legislación técnica aplicable en cada mercado y es ajena al contrato. En cuanto a la cantidad de energía de cada contrato, ésta suele ser de 1 MW de potencia durante una hora del periodo del que se trate. El punto de entrega suele ser la red de alta tensión del sistema eléctrico. El periodo de negociación también se especifica en cada mercado para cada producto.

Cuando se acerca el periodo de entrega del contrato de futuros, el precio del futuro converge hacia el precio del mercado spot o precio de contado, de forma que al llegar el periodo de entrega el precio del futuro se iguala o está muy cercano al precio de mercado. En efecto esto es así porque si el precio del futuro fuera superior al precio del mercado spot durante el periodo de entrega, esto daría lugar a una clara oportunidad de arbitraje¹ vendiendo un contrato de futuros (a precio superior al del mercado spot) y comprando en el mercado spot (a precio inferior al del contrato de futuros), con lo que se obtendría un beneficio igual a la diferencia entre el precio del futuro y el precio del mercado spot. Es claro además que a medida que los agentes exploten esta oportunidad, el precio bajará. Análogos resultados pueden obtenerse si el precio del futuro es inferior al precio spot.

La liquidación de los contratos puede realizarse por entrega (liquidación física) o sin ella (liquidación financiera).

1.2.4. Mercado a Plazos (Contratos Forward)

Los contratos a plazo son, como los contratos de futuros, un acuerdo de compra o venta de energía eléctrica a un cierto precio en una fecha futura dada.

Sin embargo, los contratos a plazo no son negociados como tal en un mercado organizado, sino que son acuerdos privados entre agentes, aunque a pesar de esto, los mercados organizados también pueden ofrecer contratos estandarizados de forwards a sus clientes. Otra diferencia importante con respecto a los contratos de futuros es que los contratos a plazo no tienen liquidación diaria entre el momento de la compra/venta y la fecha de suministro de la energía, sino que durante ese periodo se van acumulando los beneficios y pérdidas diarias y el saldo final se liquida al vencimiento, durante el periodo en el cual se ejecuta la entrega física de la energía.



1.2.5. Contratos de Opciones

Una opción es un contrato a través del cual el emisor, por una cierta cantidad de dinero llamada prima, da al comprador el derecho de exigir dentro de un plazo determinado, la compra (call) o venta (put) de un activo (en nuestro caso energía eléctrica) a un precio fijo llamado precio de ejercicio. Por lo tanto, a diferencia de un contrato de futuros, el comprador de una opción no está obligado a ejecutarla, simplemente tiene el derecho a hacerlo.

Si la opción puede ser ejercida en cualquier instante, se la llama opción americana. Si la opción sólo se puede ejercer en una fecha determinada, recibe el nombre de opción europea.

Las opciones no son un producto muy común en los mercados eléctricos.

1.2.6. Operador de Sistema

En el entorno tradicional previo a la liberalización los gobiernos de la mayoría de los países, con independencia de si los activos eléctricos eran o no de propiedad pública, realizaban la planificación centralizada de las inversiones necesarias en el sector, pudiendo ésta ser realizada por empresas públicas o privadas, mientras que las empresas monopolísticas de cada región eran las responsables de alcanzar un determinado nivel de fiabilidad en el suministro. El despacho de generación en tiempo real era habitualmente realizado por la propia empresa verticalmente integrada, siguiendo el orden de mérito en cada región. Cuando existía posibilidad de intercambio de energía entre regiones limítrofes, ésta se llevaba a cabo siguiendo acuerdos establecidos previamente. Las obligaciones del centro de control de la región eléctrica en la que operaba la empresa verticalmente integrada, que habitualmente era parte de la propia empresa, eran la seguridad del sistema, el despacho de la generación, el mantenimiento de un nivel de reserva suficiente para hacer frente a posibles contingencias en el sistema, y la gestión técnica de la capacidad y disponibilidad de la red de transporte.

En el nuevo entorno liberalizado la operación del sistema es realizada centralizadamente por el Operador del Sistema, OS, que garantiza el funcionamiento del sistema eléctrico en condiciones de seguridad y calidad compatibles con las decisiones de producción y consumo decididas por los agentes del mercado bien como resultado de la casación de ofertas llevada a cabo por el OM o bien mediante contratos bilaterales, sustituyendo así a los tradicionales procedimientos de minimización de los costes de producción. Otras funciones habituales del OS son la gestión de los servicios complementarios, el diseño de los procedimientos de operación, y la planificación a largo plazo.

Por su naturaleza, la remuneración de la actividad de operación del sistema debe estar sujeta a regulación. Además, se debe garantizar la independencia del OS para asegurar que el desempeño de sus funciones se realiza sin que exista un trato discriminatorio a los agentes.



1.3. ALGUNOS MERCADOS ELÉCTRICOS EUROPEOS

1.3.1. España

A lo largo de estos últimos años, desde 1998 hasta la actualidad, se han llevado a cabo muchos cambios regulatorios y legislativos con el fin de dotar al mercado de las características típicas y esperadas de un mercado eléctrico liberalizado como, por ejemplo, justa competencia o liquidez. El patrón a seguir es el NordPool escandinavo, donde se ha llegado a negociar un volumen equivalente a siete veces la electricidad que se acabó consumiendo o al EEx (European Energy Exchange) donde se ha negociado cinco veces la producción real.

Se puede remarcar que la liberalización supuso unos cuantos acontecimientos como:

- Libre creación y entrada de nuevas empresas en el sector, tanto nacionales como extranjeras.
- Posibilidad de llevar a cabo transacciones bilaterales entre demandantes y oferentes.
- Creación del Operador del Mercado con la misión de gestionar económicamente el mercado. OMEL (Operador del Mercado Eléctrico Español) es quien tiene esta función.
- Creación del Operador del Sistema con la responsabilidad de la gestión técnica de todo el sistema eléctrico. Es REE quien se ocupa de mantener en buen estado la red de transporte y de alta tensión.

Red Eléctrica opera el sistema eléctrico español, tanto en la península como en los sistemas insulares y extrapeninsulares, garantizando la seguridad y continuidad del suministro eléctrico.

Es Red Eléctrica quien elabora anualmente las previsiones de evolución de la demanda eléctrica a medio y largo plazo, así como de su cobertura. Estas previsiones son fundamentales para la elaboración de los planes de desarrollo de la red de transporte para los próximos años, aprobados por el Ministerio de Industria, Turismo y Comercio.

Red Eléctrica también gestiona los denominados servicios de ajuste que son aquellos que tienen por objeto adecuar los programas de producción resultantes de los mercados eléctricos diarios e intradiarios a los requisitos de calidad, fiabilidad y seguridad del sistema eléctrico. Se entienden por servicios de ajuste o mercados de ajuste la solución de restricciones técnicas, la asignación de los servicios complementarios y la gestión de desvíos.

1.3.2. Austria

En Austria la liberalización energética se ha orientado hacia la liberalización de la generación y el suministro, mientras que la distribución y transmisión de electricidad se mantuvo bajo control estatal. El Verbundgesellschaft (Austria posee el 51% de esta



empresa) conserva su papel fundamental de la mayor productora y transportista nacional de energía.

La regulación del mercado de la energía de Austria se lleva a cabo a través de dos organismos públicos. El primero es el Energie-Control Kommission (E-Comisión de Control) que tiene el poder legislativo y también poder judicial. El segundo, Energie-Control GmbH (E-Control), actúa como un regulador de la competencia y de los precios de la energía. Austria inició una apertura gradual del mercado en 1999, que fue completado en su totalidad en febrero de 2003.

1.3.3. Italia

En Italia antes de la liberalización, el mercado eléctrico italiano fue tradicionalmente dominado por la empresa monopolio de Ente Nazionale per l'Energia Elettrica (ENEL). La aplicación de la normativa sobre la electricidad de la UE en primer lugar, inició la apertura del mercado eléctrico italiano a la competencia, que se ha logrado mediante la privatización de ENEL. Un nuevo órgano estatal, l'Autorità per l'Energia Elettrica e il gas (AEEG), tiene como principales tareas regular, controlar y supervisar el sector de electricidad y de gas en Italia.

Se ha decidido preservar como propiedad exclusiva del Estado el suministro de electricidad. Asimismo, el Estado conserva en propiedad exclusiva el transporte de energía eléctrica hasta el año 2005. Esto se logró con la iniciativa impulsada por el estado en 1999, que creó el Transmission System Operator italiana (GRTN, Gestore della Rete di trasmissione Nazionale SpA), controlada por el Ministerio de Economía y Finanzas. En 2005 esta actividad fue transferida a Terna SpA. En la actualidad, además de ser responsable de las habituales actividades del GRT: la transmisión de electricidad, el envío, el suministro de electricidad seguridad, la creación de la reservas de energías, la red de mantenimiento y servicios de equilibrio en toda Italia, Terna SpA es dueño de casi toda la electricidad de alta tensión de Italia.

1.3.4. Francia

La liberalización del sector eléctrico francés se produce como consecuencia de la transposición de la directiva europea de 1996 en la ley francesa de 10 de febrero de 2000. El análisis para la implantación de un nuevo mercado fue llevado a cabo por EURONEXT PARIS con la participación de diversos agentes energéticos europeos y el comienzo del mercado organizado tuvo lugar en Noviembre de 2001.

RTE (Electricity Transmission Grid) es una entidad estatal responsable de mantener la fiabilidad y seguridad del sistema, asegurando que se realizan las transferencias energéticas acordadas en el mercado. RTE posee y opera la red de transporte desde 63 kV a 400kV, siendo una entidad englobada en EDF, aunque con actividades separadas del resto:

- Gestión independiente. El director es nombrado por el gobierno con la opinión de la Comisión de Regulación.



- Separación contable.
- Confidencialidad e imparcialidad definidas por ley. Su remuneración se realiza mediante tarifa.

POWERNEXT S.A. es el único operador del mercado francés con un capital inicial de 10 millones de euros. Su accionariado está compuesto por sociedades tanto públicas como privadas. Entre ellos cabe destacar HGRT, un consorcio de operadores europeos de transporte.

Los reguladores son:

- Banque de France, como supervisor de los mercados de capital y las instituciones financieras.
- Conseil des Marchés Financiers (CMF), como encargado de la supervisión y regulación de los mercados regulados y OTC. Tiene capacidad sancionadora.
- Comisión des Opérations de Bourse (COB). Supervisa la información suministrada a los inversores en las ofertas públicas y la viabilidad financiera de los contratos de los diversos agentes.
- Comisión de Régulation de l'Electricité (CRE). Se encarga de la regulación del mercado de electricidad, así como del acceso a las redes de transporte y distribución.
- Direction du Trésor. Propone el nombramiento de los miembros de CMF y valida sus normas y regulaciones. Sus decretos son firmados por el Ministro de Economía y Finanzas.

1.3.5. Nordpool

La liberalización de la electricidad en la región nórdica se puede clasificar como una iniciativa regional que, a diferencia de otros países europeos, cronológicamente es anterior a la Primera Normativa europea sobre la electricidad y no ha sido motivado por la misma. Historicamente el primer intercambio de energía en Nordpool fue una iniciativa Noruega en 1993 y fue gestionado por una entidad vinculada a la empresa nacional de TSO Statnett. En 1996, Suecia liberalizó su mercado de energía e hizo un intercambio conjunto, Nord Pool, con Noruega. Más tarde, en 1996, Finlandia y Dinamarca, en 1999-2000 también se unieron a la iniciativa. En el año 2002 los mercados financieros de Nord Pool se dividieron. Como consecuencia, una nueva empresa, Spot Nord Pool AS, ha sido creada para administrar el mercado spot de Noruega, Dinamarca, Finlandia, Suecia y más recientemente la región Kontek en Alemania. La propiedad de la nueva compañía se divide entre los GRT de Noruega, Dinamarca, Finlandia y Suecia. La Dirección de Recursos Hídricos de Noruega y de Energía supervisa y regula la actividad del mercado spot de Nord Pool. La Comisión de Valores de Banca y Seguros supervisa el mercado.



1.3.6. Alemania

La implantación de la directiva IEM en Alemania se realizó principalmente a través de la nueva ley de energía revisada (“Gesetz zur Neuregelung des Energiewirtschaftsrechts”), que entró en vigor el 29 de Abril de 1998 declarando con capacidad de elección al 100% de los consumidores directamente.

En la primera mitad de 2000 se establece la bolsa LPX (Leipzig Power Exchange GmbH) como mercado de electricidad alemán en el que la sociedad responsable facilita a la bolsa de electricidad de Leipzig (LPX) los medios materiales, personales y financieros. LPX es socio financiero en el negocio al contado y asume la ejecución financiera de los negocios. En 2001 LPX lanza un mercado financiero de futuros y opciones.

El 8 de agosto de 2000 comienza sus actividades EEX European Energy Exchange, con sede en Frankfurt y un mercado spot, el EEX Spot Market. El 1 de marzo de 2001 EEX lanza su mercado de futuros, el EEX Derivatives market.

El 1 de marzo de 2002 los dos Power Exchanges alemanes se fusionan con efectos retroactivos al 1 de enero. En el mercado resultante de la fusión se incorpora también un área de servicios de clearing para operaciones OTC, siempre y cuando los instrumentos negociados sean idénticos a los negociados en el mercado de futuros. El nuevo mercado actuará como cámara de compensación en todas las operaciones en las que intermedie.

En Alemania no existe como tal un único Operador del Sistema, sino que los propietarios de las redes de transporte las operan de forma descentralizada. La planificación y operación de la red de transporte es responsabilidad de los propietarios de las redes. Además, el acceso a la red es negociado, lo cual, en la práctica, supone un límite a la competencia.

Como forma de colaboración, el 15 de diciembre de 1948 se fundó la DVG Deutsche Verbundgesellschaft, como asociación de los transportistas alemanes con responsabilidades sobre la seguridad del sistema, para combinar la organización descentralizada con las ventajas de la actividad conjunta y la cooperación entre los distintos miembros, pero manteniendo la independencia de los mismos.



- | | |
|--------------------------|---------------------------------------|
| 1 Bewag AG | 4 Hamburgische Electricitäts-Werke AG |
| 2 EnBW Transportnetze AG | 5 RWE Net AG |
| 3 E.ON Netz GmbH | 6 VEAG Vereinigte Energiewerke AG |

Al final de 2001, la DVG cesa sus actividades y los intereses de los operadores alemanes, tanto de transporte como de redes de menor tensión, se concentran en “Verband der Netzbetreiber - VDN - e.V. beim VDEW “, que había comenzado sus actividades en el verano de 2001. Como se ha dicho, no existe la figura del Operador del Sistema.

El Ministerio de Economía es la autoridad general en el contexto de la ley alemana de energía. Las autoridades federales y municipales son las responsables de resolver los conflictos de acceso a la red o los relacionados con el incumplimiento de las reglas de la competencia, como el abuso de posiciones dominantes. Las bolsas son entidades de derecho público con capacidad jurídica. El control de la bolsa es asunto de los estados federados y en el estado de Sajonia es asumido por el Ministerio Estatal Sajón para la Economía y el Trabajo.



1.4. MERCADOS ELÉCTRICOS INTERNACIONALES

1.4.1. Ontario

Antes de la liberalización, Ontario Hydro junto con algunas pequeñas empresas de servicios públicos municipales, generaba, transmitía y distribuía la electricidad a sus clientes en toda la provincia. En esa época, los precios de la electricidad estaban regulados por el gobierno de la provincia. En 1998, el gobierno de Mike Harris aprobó la ley de competencia de Energía que autorizó el establecimiento de un mercado de la electricidad. En abril de 1999, Ontario Hydro fue re-organizada en cinco empresas: el Operador de Mercado Independiente (IMO), Hidro One, Ontario Power Generation Inc. (OPG), la Autoridad de Seguridad Eléctrica (ESA) y la Sociedad Financiera de la Electricidad de Ontario (OEFC). La ESA es responsable de los estándares de la industria eléctrica, y el OEFC gestiona los servicios financieros de la antigua Ontario Hydro y sus sucesores.

El mercado mayorista de electricidad de Ontario abrió el 1 de mayo de 2002, dos años después de la fecha originalmente programada. Este mercado se compone de un mercado en tiempo real de la energía y de explotación de reservas, y un mercado financiero. La Ley de Electricidad de 2004, renombró la IMO como el Operador Independiente del Sistema Eléctrico (IESO). La IESO es una empresa sin ánimo de lucro que es regulada por la Junta de Energía de Ontario y su responsabilidad principal es operar en el mercado mayorista de electricidad de Ontario.

El sistema de distribución está regulada también por la Junta de Energía de Ontario con 91 empresas locales de distribución que suministran electricidad a clientes al por menor. El sistema de distribución de alta tensión de Ontario está conectado con Manitoba, Quebec, Nueva York, Michigan y Minnesota a través de 12 líneas.

Ontario Power Authority fue creada en 2005, bajo las instrucciones de la Ley de la Electricidad de 2004, para garantizar la fiabilidad, la seguridad y la eficiencia del sector eléctrico de Ontario.

La oferta y la demanda dependen de entidades dentro de la provincia que tienen una relación directa con la red de distribución y que deben participar en el mercado eléctrico de Ontario. Este grupo de entidades está formado por las empresas de generación y de distribución. Otras entidades que están conectados a la red de distribución se les conocen como instalaciones "embedded" y pueden optar por participar ya sea en el mercado o en la compra / venta de energía a través de contratos con los minoristas de energía. Hay, sin embargo, otros participantes del mercado sin una conexión física, como son los comerciantes de energía, o entidades que importan / exportan energía a / de Ontario y pueden participar en el mercado.

Los participantes en el mercado de energía en Ontario pueden optar por comprar o vender energía mediante contratos bilaterales. Sin embargo, con los contratos bilaterales



no es necesario informar a IESO. Los contratos bilaterales no son considerados en el proceso de programación de la energía, y tienen una pequeña participación en el comercio eléctrico de Ontario.

1.4.2. PJM

Uno de los mercados más importantes dentro de Estados Unidos es PJM, el cual existe desde 1927, cuando tres empresas eléctricas realizaron el primer acuerdo de intercambio de energía, el PA-NJ Agreement. Este acuerdo se mantuvo en vigor durante casi 30 años, hasta que PJM se expandió a su actual zona en 1956. Durante varios años antes de 1927, las empresas ya habían obtenido beneficios interconectando instalaciones de sus respectivos sistemas. Esto les llevaba a reducir las necesidades de capacidad instalada y de reserva, menores costes de operación, y a la mejora de la fiabilidad del suministro. El acuerdo hacía que cada empresa fuera la responsable técnica y económica de su red de transporte, que debía estar disponible para el intercambio de potencia entre ellas, y permitía el apoyo mutuo ante contingencias.

El 29 de Marzo de 1995 la Federal Energy Regulatory Commission (FERC) publicó una propuesta para la transición a la competencia de los mercados de electricidad. El 24 de abril de 1996 la FERC publica la Order 888 con los objetivos de eliminar las barreras a la competencia en los mercados eléctricos. Entre otras medidas, la Order 888 impuso la obligatoriedad, para los sistemas eléctricos interconectados como PJM, de crear un Operador de Sistema Independiente (Independent System Operator, ISO) encargado de la operación de la red de transporte y de la gestión de un mercado eléctrico regional. Para llevar a cabo la reestructuración PJM se sirvió de una serie de acuerdos, pero el plan fue rechazado por la FERC en primera instancia, para ser luego aceptado el 28 de Febrero de 1997, tras una serie de cambios. El 31 de marzo de 1997 los miembros de PJM deciden transformar PJM Interconnection Association en sociedad anónima, pasando a ser PJM Interconnection, LLC y el 1 de abril de ese año, PJM firma con la FERC el Operating Agreement, que sustituye al PJM Agreement de 1956. El 25 de noviembre de 1997 PJM es constituido como ISO (Independent System Operator).

El 13 de Marzo de 1999 la FERC emite una propuesta para unir todas las redes de transporte de EEUU en Regional Transmission Organizations (RTOs) y el 20 de Diciembre la Orden 2000, que obliga a las compañías propietarias de las infraestructuras de red que participan en el comercio eléctrico entre diferentes estados a formar RTOs. El 11 de octubre de 2000, PJM crea su RTO, de acuerdo con la Orden 2000.

PJM es una entidad privada encargada de la operación del sistema y la operación del mercado eléctrico. También es el operador de la red de transporte (TSO, Transmission System Operator), y mantiene acuerdos con los propietarios de la red, cuya propiedad es independiente de PJM.

PJM es el Operador del Sistema, independiente de las actividades de la red de transporte.

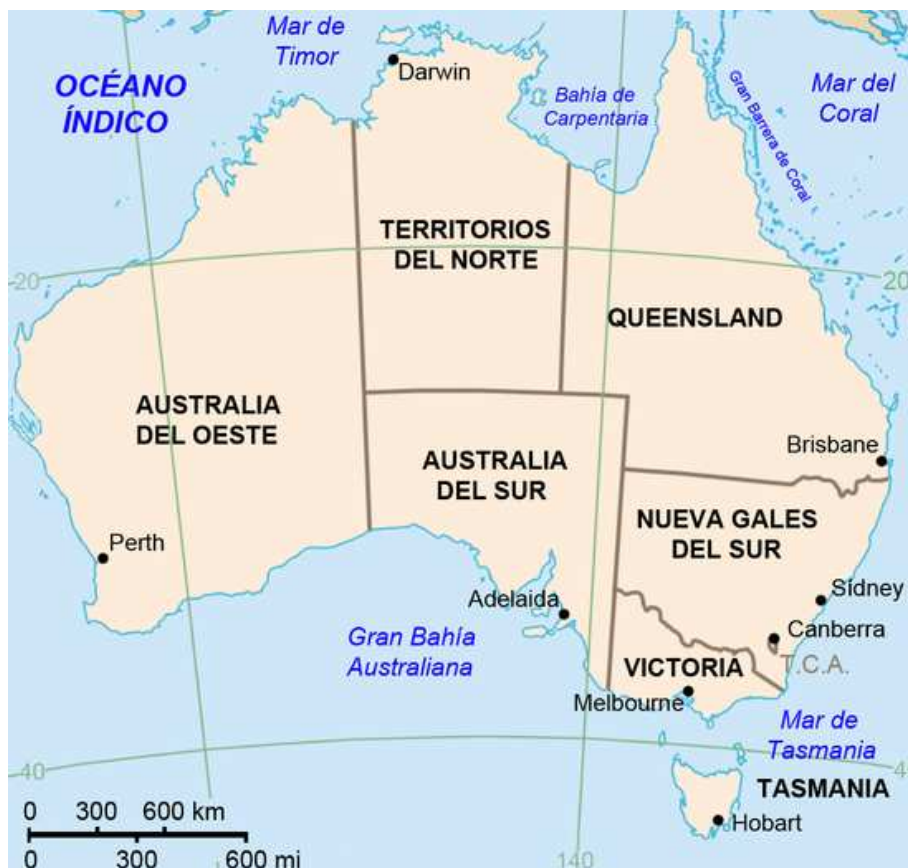


En Pennsylvania, el regulador es Pennsylvania PUC. Además, FERC es la entidad que define las normas de funcionamiento de PJM.

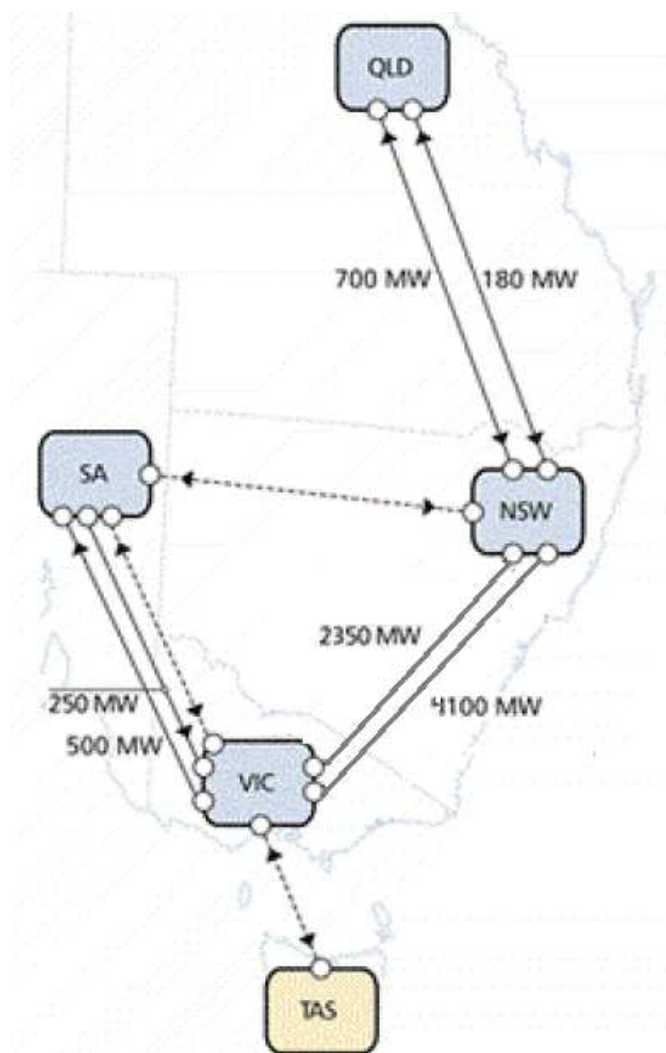
1.4.3. Australia

El sector eléctrico australiano está fuertemente influenciado por su estructura territorial. Australia está compuesto por:

- Seis estados federales:
 - Victoria, Nueva Gales del Sur y Queensland, en los que se concentra la mayor parte de la población y, por lo tanto, de la demanda eléctrica.
 - South Australia (también conocida como Australia Meridional), Western Australia y Tasmania, mucho menos poblados.
- Dos “mainland territories”: Australian Capital Territory y Northern Territory.



Como resultado de esta estructura territorial, el sector eléctrico se caracteriza por mercados de carácter regional e interconexiones de capacidad limitada, encontrándose Tasmania, Western Australia y Northern Territory aislados eléctricamente de los estados del este:



Antes de la implantación de un mercado competitivo, el suministro eléctrico era tradicionalmente responsabilidad de cada Estado a través de empresas públicas verticalmente integradas que actuaban en régimen de monopolio.

En 1991 se creó The National Grid Management Council (NGMC), un órgano consultivo del gobierno para la introducción de competencia en el sur y este de Australia (Tasmania, Western Australia y Northern Territory quedaron fuera de estas reformas, por estar, como se ha dicho, aislados eléctricamente del este). Sin embargo, aunque la decisión de emprender la reforma del sector ya había sido tomada, el mercado eléctrico nacional, National Electricity Market (NEM), no entró en operación hasta el 13 de diciembre de 1998.

En paralelo a este proceso nacional, que duró ocho años, los distintos estados llevaron a cabo sus propias reformas. Por ejemplo, los estados de Nueva Gales del Sur y Victoria introdujeron sus propios mercados organizados antes de entrar en vigor el NEM, en el que fueron posteriormente unificados en 1997.

En general, Australia considera que la competencia total en el sector eléctrico puede ser alcanzada sin necesidad de la privatización completa de los activos de las actividades a



desarrollar en competencia: la decisión del tipo de propiedad depende de cada estado y sólo Victoria y Nueva Gales del Sur han optado por la privatización total.

El Operador del Sistema y el Operador del Mercado son una única entidad, NEMMCO, que es propiedad de los cinco gobiernos estatales, correspondientes, aproximadamente, a las cinco regiones eléctricas interconectadas que participan en el mercado: New South Wales (que incluye Australian Capital Territory), Queensland, South Australia, Victoria y Snowy. Cada Gobierno aporta un miembro a la junta directiva. La red de transporte es única, aunque existen Operadores Locales de Red, coordinados por el Operador del Sistema.

La regulación del sector eléctrico se encuentra bajo jurisdicción estatal, con cierta responsabilidad del Gobierno de la Commonwealth. Por lo tanto, cualquier reforma a escala nacional requiere un acuerdo entre los estados y la Commonwealth. “The Snowy Mountains Hydro-electric Scheme” es un proyecto llevado a cabo conjuntamente por los gobiernos de Nueva Gales del Sur, Victoria y la Commonwealth. La generación y el transporte son responsabilidad de Snowy Mountains Hydro-electric Authority y la venta al mercado de la energía es llevada a cabo por la compañía pública Snowy Hydro Trading Pty Ltd.

Existe un Regulador Federal, The National Electricity Code Administrator Limited, NECA, y reguladores Estatales, dependiendo ambos del Gobierno Nacional. Además, The Australian Competition and Consumer Commission, ACCC, determina la remuneración del servicio de transporte y aprueba los cambios en The National Electricity Code. Por otro lado, The Australian Securities and Investment Commission, ASIC, es la encargada de regular los instrumentos financieros utilizados por los agentes para la cobertura del riesgo derivado de su participación en el mercado.



2. CÓMO DETECTAR DEPENDENCIAS ENTRE SERIES USANDO PERMUTACIONES

Un problema frecuente en el contexto de las series temporales consiste en determinar si dos series de datos (x_t) e (y_t) son independientes o no. Con tal fin, se han desarrollado diferentes tests estadísticos que han permitido resolver el problema, aunque sólo en parte. Algunos ejemplos son los trabajos de Haugh (1976), Pierce (1977), Geweke (1981) y Hong (1996), con la limitación de detectar sólo relaciones lineales a través de las correlaciones cruzadas. Recientemente, Matilla et al (2010) y más tarde Cánovas et al (2011) han propuesto tests de independencia entre series basados en cálculo simbólico, permitiendo detectar también relaciones de tipo no lineal.

En el contexto de los mercados energéticos, podemos encontrar algunos estudios recientes en los que se analizan dependencias entre series de interés, y que se describen a continuación.

Alexander (2005) estudió la dependencia entre los precios del crudo y gas natural, concluyendo que la dependencia es fuerte, de tipo no lineal, y que no se puede modelar correctamente por una distribución normal bivariada. Es más, cuando los precios de estos productos se ponen como un modelo de serie temporal, las distribuciones para los términos de error (residuos) resultan ser asimétricos.

El trabajo de Grégoire et al (2008) muestra cómo las predicciones para los precios del crudo y del gas natural se pueden mejorar modelizando previamente la dependencia existente entre ambas series, dependencia que es capturada mediante el uso de Mihaylova (2009) analizó las dependencias entre precios de la electricidad por hora, en base a cinco conjuntos de datos históricos de energía, correspondientes a Austria, Alemania, Nord Pool, España y Suiza. En este caso, el estudio de la dependencia existente entre los precios de la energía de los países seleccionados se realizó mediante el uso de cópulas y dos tipos de coeficientes de correlación (tau de Kendall y rho de Spearman). Entre sus conclusiones podemos destacar que se observó la mayor dependencia entre los precios de la electricidad de Austria y Alemania, mientras que la dependencia más débil se presentó entre el Nord Pool y España.

Por otra parte, Erni (2009) investigó si los mercados de la electricidad de Suiza, Austria y Alemania se integran y convergen hacia un solo precio. Basado en un análisis de cointegración y una aplicación del Filtro de Kalman, se analizaron esos tres mercados desde 2007 a 2009. El estudio concluyó una estrecha relación entre los precios del mercado alemán y del austriaco, y a su vez con el mercado suizo aunque de forma más débil. Además, se comprobó que durante los periodos de verano los tres precios están muy cointegrados y se aplica la ley del precio único.



Basándonos en la metodología empleada en el trabajo de Cánovas et al (2011), en este proyecto proponemos estudiar las dependencias existentes entre algunos mercados eléctricos mediante una codificación previa de la serie a través de las permutaciones (cálculo simbólico y entropía de permutación).

La primera sección de este capítulo está dedicada a introducir el proceso de codificación de una serie y definir el concepto de entropía de permutación, así como algunas de sus aplicaciones. En la segunda sección, se describe la utilidad de las tablas de contingencia para detectar asociación o dependencia entre variables y se introducen algunas medidas de asociación entre variables. La sección tres está dedicada a introducir nuestra propuesta de medidas de asociación y de “distancia” entre dos series temporales, mediante el uso de las permutaciones.

2.1. ENTROPÍA DE PERMUTACIÓN

La entropía de permutación, introducida por Bandt y Pompe (2002), se utiliza para estudiar la complejidad presente en una serie temporal. De forma resumida podemos decir que consta de dos etapas: una primera etapa de codificación de la serie mediante el uso de permutaciones y una segunda etapa de cálculo de la entropía de Shannon para la distribución discreta resultante de la codificación.

Etapas 1 (Codificación mediante permutaciones):

Para un entero positivo $m \geq 2$, que denominaremos dimensión *embedding*, consideremos el grupo con cardinal $m!$ formado por todas las permutaciones de longitud m :

$$S_m = \{\pi_1, \pi_2, \dots, \pi_{m!}\}$$

donde cada elemento π , que denominaremos símbolo, es de la forma:

$$\pi = (i_1, i_2, \dots, i_m) \in S_m$$

Una forma natural de codificar una serie temporal es a través de las permutaciones que aparecen a lo largo de dicha serie, es decir, teniendo en cuenta sólo el orden existente entre los datos de la serie y no su valor numérico.

Sea:

$$(x_n)_{n=1}^T$$

una serie temporal unidimensional real de longitud T y sea

$$x_m(r) = (x_r, x_{r+1}, \dots, x_{r+m-1}), 1 \leq r < T - m + 1$$

una ventana móvil de tamaño m extraída de la serie anterior. Diremos que dicha ventana móvil es de tipo:



$$\pi = (i_1, i_2, \dots, i_m) \in S_m.$$

si y sólo si π es el único elemento de S_m verificando las siguientes condiciones:

$$(c1) x_{r+i_1} \leq x_{r+i_2} \leq \dots x_{r+i_m}.$$

$$(c2) i_{s-1} \leq i_s \text{ if } x_{r+i_{s-1}} = x_{r+i_s}.$$

Observar que la condición (c2) garantiza la unicidad en la asignación del símbolo aún en el caso de valores iguales.

Por tanto, para una serie real unidimensional $(x_n)_{n=1}^T$ y fijado un embedding m , a cada ventana móvil de longitud m se le asigna un símbolo que se corresponde con la permutación que forman los valores de dicha ventana móvil.

Ejemplo ilustrativo: Consideremos la siguiente serie temporal real formada por 10 datos:

$$(x_n)_{n=1}^{10} = \{7, 5, 2, 8, 3, 9, 6, 2, 4, 1\}$$

y fijemos embedding $m=3$, de manera que el grupo de las permutaciones de orden 3 está formado por las siguientes 6 permutaciones distintas de longitud 3:

$$S_3 = \{ \pi_1 = (1,2,3), \pi_2 = (1,3,2), \pi_3 = (2,1,3), \pi_4 = (2,3,1), \pi_5 = (3,1,2), \pi_6 = (3,2,1) \}$$

Para la serie en estudio, se pueden extraer un total de $T-m+1 = 10-3+1 = 8$ ventanas móviles, cuyos símbolos asociados son los siguientes:

- Ventana móvil 1: $x_3(1)=(7,5,2) \rightarrow \text{símbolo asociado} = (3,2,1) = \pi_6$
- Ventana móvil 2: $x_3(2)=(5,2,8) \rightarrow \text{símbolo asociado} = (2,1,3) = \pi_3$
- Ventana móvil 3: $x_3(3)=(2,8,3) \rightarrow \text{símbolo asociado} = (1,3,2) = \pi_2$
- Ventana móvil 4: $x_3(4)=(8,3,9) \rightarrow \text{símbolo asociado} = (2,1,3) = \pi_3$
- Ventana móvil 5: $x_3(5)=(3,9,6) \rightarrow \text{símbolo asociado} = (1,3,2) = \pi_2$
- Ventana móvil 6: $x_3(6)=(9,6,2) \rightarrow \text{símbolo asociado} = (3,2,1) = \pi_6$
- Ventana móvil 7: $x_3(7)=(6,2,4) \rightarrow \text{símbolo asociado} = (3,1,2) = \pi_5$
- Ventana móvil 8: $x_3(8)=(2,4,1) \rightarrow \text{símbolo asociado} = (2,3,1) = \pi_4$

Etapla 2 (Cálculo de la entropía de Shannon):

Dada una función puntual de probabilidad discreta $\{p_1, p_2, \dots, p_k\}$ asociada a una variable aleatoria X , la entropía de Shannon de X se define como sigue:

$$H(X) = - \sum p_i \log(p_i)$$

Para la variable discreta resultante de la codificación de la etapa anterior, podemos obtener su función puntual de probabilidad de forma sencilla, simplemente calculando



la frecuencia de aparición de cada símbolo (permutación) y dividiéndola entre el número total de ventanas móviles extraídas de la serie.

Así, para cada símbolo o permutación π_i , la probabilidad asociada a dicho símbolo es:

$$p_{\pi_i} = p(\pi_i) = (\text{número de ventanas móviles de tipo } \pi_i) / T-m+1$$

Aplicando la entropía de Shannon a la función de probabilidad discreta resultante, obtenemos la definición de **entropía de permutación** para la serie $(x_n)_{n=1}^T$:

$$H(x_n, m) = - \sum_{\pi_i \in S_m} (p_{\pi_i} * \log(p_{\pi_i}))$$

Ejemplo ilustrativo (continuación): Continuando con el ejemplo ilustrativo de la etapa anterior, a continuación se muestran las probabilidades asociadas a cada símbolo de la serie.

- $p_{\pi_1} = p(\pi_1) = (\text{número de ventanas móviles de tipo } \pi_1) / 8 = 0 / 8 = 0$
- $p_{\pi_2} = p(\pi_2) = (\text{número de ventanas móviles de tipo } \pi_2) / 8 = 2 / 8 = 1/4$
- $p_{\pi_3} = p(\pi_3) = (\text{número de ventanas móviles de tipo } \pi_3) / 8 = 2 / 8 = 1/4$
- $p_{\pi_4} = p(\pi_4) = (\text{número de ventanas móviles de tipo } \pi_4) / 8 = 1 / 8 = 1/8$
- $p_{\pi_5} = p(\pi_5) = (\text{número de ventanas móviles de tipo } \pi_5) / 8 = 1 / 8 = 1/8$
- $p_{\pi_6} = p(\pi_6) = (\text{número de ventanas móviles de tipo } \pi_6) / 8 = 2 / 8 = 1/4$

Por tanto, la entropía de permutación de la serie del ejemplo viene dada por:

$$H(x_n, m) = - \sum_{\pi_i \in S_m} (p_{\pi_i} * \log(p_{\pi_i})) =$$

$$-(0 + 3 * 1/4 * \log(1/4) + 2 * 1/8 * \log(1/8)) = 0'6773$$

Como el objetivo de este proyecto es estudiar relaciones entre pares de series temporales, a continuación se describe cómo se desarrolla la etapa de codificación en el caso de series temporales bidimensionales.

Codificación para series temporales bidimensionales:

Consideremos dos series temporales del mismo tamaño, (x_n) e (y_n) , y la serie bidimensional formada por las dos anteriores (w_n) , con $w_i = (x_i, y_i)$. Fijemos un embedding m y consideremos la ventana móvil bidimensional $w_m(r) = (x_m(r), y_m(r))$. Diremos que la ventana móvil bidimensional $w_m(r)$ es de tipo $\eta_{ij} = (\pi_i \times \pi_j) \in S_m \times S_m$ si y sólo si $x_m(r)$ es de tipo π_i e $y_m(r)$ es de tipo π_j . A η_{ij} lo llamaremos símbolo de la serie bidimensional.

Igual que en el caso unidimensional, podemos determinar la probabilidad asociada a cada símbolo de la serie bidimensional del siguiente modo:

$$p(\eta_{ij}) = p(\pi_i \times \pi_j) = (\text{número de ventanas móviles de tipo } \pi_i \times \pi_j) / T-m+1$$



y cuyas probabilidades marginales se corresponden con las funciones discretas de probabilidad de la codificación de (x_n) e (y_n) , respectivamente.

Ejemplo ilustrativo: Consideremos la serie temporal $(x_n)_{n=1}^{10}$ del ejemplo desarrollado más arriba y esta otra serie del mismo tamaño:

$$(y_n)_{n=1}^{10} = \{3, 8, 2, 5, 1, 4, 7, 6, 3, 9\}$$

así como la serie bidimensional (w_n) , con:

$$w_i = (x_i, y_i), i=1, \dots, 10.$$

Fijemos embedding $m=3$, de manera que el grupo de las permutaciones de orden 3 está formado por las siguientes 6 permutaciones distintas de longitud 3:

$$S_3 = \{ \pi_1 = (1,2,3), \pi_2 = (1,3,2), \pi_3 = (2,1,3), \pi_4 = (2,3,1), \pi_5 = (3,1,2), \pi_6 = (3,2,1) \}$$

Para la serie bidimensional en estudio, se pueden extraer un total de $T-m+1 = 10-3+1 = 8$ ventanas móviles bidimensionales, cuyos símbolos asociados son los siguientes:

- Ventana móvil 1: $w_3(1) = [(7,5,2), (3,8,2)] \rightarrow \text{símbolo asociado} = \pi_6 \times \pi_4$
- Ventana móvil 2: $w_3(2) = [(5,2,8), (8,2,5)] \rightarrow \text{símbolo asociado} = \pi_3 \times \pi_5$
- Ventana móvil 3: $w_3(3) = [(2,8,3), (2,5,1)] \rightarrow \text{símbolo asociado} = \pi_2 \times \pi_4$
- Ventana móvil 4: $w_3(4) = [(8,3,9), (5,1,4)] \rightarrow \text{símbolo asociado} = \pi_3 \times \pi_5$
- Ventana móvil 5: $w_3(5) = [(3,9,6), (1,4,7)] \rightarrow \text{símbolo asociado} = \pi_2 \times \pi_1$
- Ventana móvil 6: $w_3(6) = [(9,6,2), (4,7,6)] \rightarrow \text{símbolo asociado} = \pi_6 \times \pi_2$
- Ventana móvil 7: $w_3(7) = [(6,2,4), (7,6,3)] \rightarrow \text{símbolo asociado} = \pi_5 \times \pi_6$
- Ventana móvil 8: $w_3(8) = [(2,4,1), (6,3,9)] \rightarrow \text{símbolo asociado} = \pi_4 \times \pi_3$

Y las probabilidades asociadas a cada símbolo de la serie son:

- $p(\pi_6 \times \pi_4) = 1/8$
- $p(\pi_3 \times \pi_5) = 2/8 = 1/4$
- $p(\pi_2 \times \pi_4) = 1/8$
- $p(\pi_2 \times \pi_1) = 1/8$
- $p(\pi_6 \times \pi_2) = 1/8$
- $p(\pi_5 \times \pi_6) = 1/8$
- $p(\pi_4 \times \pi_3) = 1/8$
- $p(\pi_i \times \pi_j) = 0$, en el resto de casos

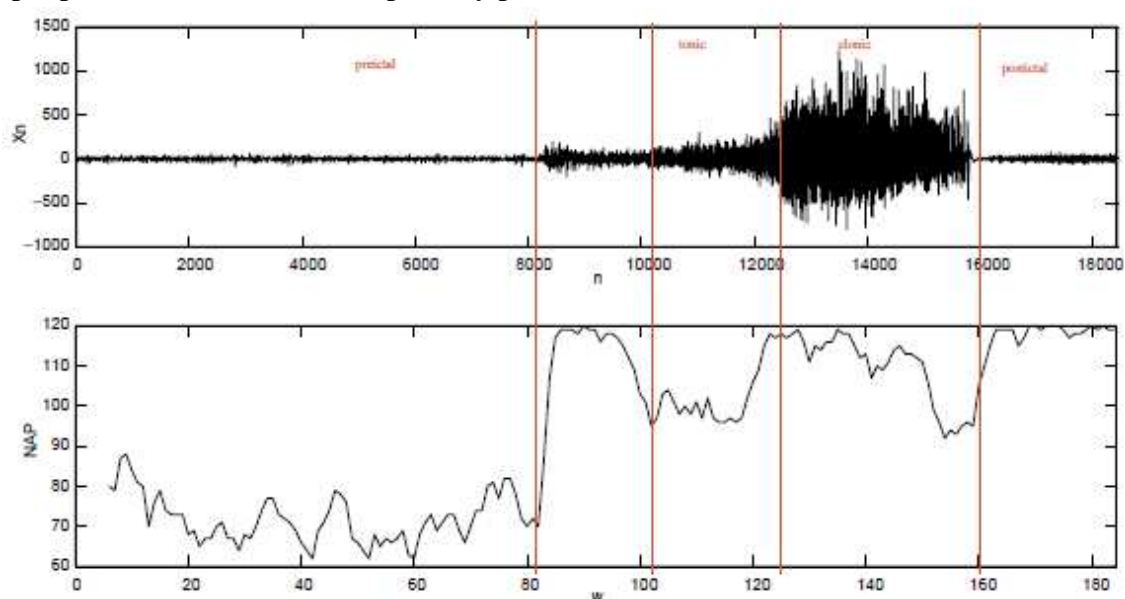
Algunas aplicaciones de la Entropía de Permutación

Una de las principales utilidades de la entropía de permutación es el estudio de la complejidad de una serie, así como la detección de cambios dinámicos en la misma.

El problema de la detección de cambios dinámicos en sistemas complejos aparece en distintas disciplinas como la Física, Medicina, Ingeniería y Ciencias Económicas. Así, en meteorología es importante detectar los instantes y lugares de cambio en el clima, o en medicina, una adecuada detección de la transición entre estados normal y anormal puede mejorar el diagnóstico y el tratamiento. Algunas referencias interesantes son Bandt y Pompe (2002), Cao et al (2004), Olofsen et al (2008) y Cánovas et al (2011). Destacar que en este último trabajo, Cánovas et al (2011) muestran que en realidad, para detectar los cambios de complejidad de una serie, basta con calcular el número de permutaciones distintas que aparecen a lo largo de la serie, proporcionando resultados muy similares a la aplicación de la entropía de permutación.

Veamos un ejemplo concreto de la utilidad de la entropía de permutación, aplicado en este caso a un tipo de ataque epiléptico denominado tónico-clónico. Una convulsión tónico clónica es un tipo de ataque de epilepsia que afecta al cerebro y que se divide en dos fases, la tónica y las clónicas. Debido al agotamiento físico y nervioso, un sueño post-ictal siempre sigue a una convulsión tónico-clónica. La confusión y amnesia total son normales tras recobrar la conciencia y poco a poco desaparece a medida que la persona es consciente del ataque sufrido.

La siguiente figura muestra una señal de electroencefalograma (EEG) correspondiente a un ataque tónico-clónico, y recogida en la zona craneal central derecha de un individuo. Contiene un total de tres minutos de grabación, con alrededor de un minuto previo a las convulsiones, las convulsiones y alguna actividad posterior al ataque. Además de la señal EEG, la figura muestra en la parte inferior el número de permutaciones distintas que aparecen a lo largo de la serie temporal, usando un embedding de $m = 5$ y un tamaño de ventana móvil $w = 600$ datos. El gráfico muestra la eficacia de las permutaciones para detectar con claridad las distintas fases del ataque epiléptico, así como las zonas previa y posterior a la convulsión.





Como la señal EEG se registró con un electrodo del cuero cabelludo, la actividad muscular rápidamente impide ver claramente la grabación, haciendo que el análisis directo de la señal EEG sea más complicado. Sin embargo, se puede comprobar la eficacia del número de permutaciones para detectar el inicio de la crisis epiléptica, las convulsiones y sus diferentes fases incluso en presencia artefactos que producen algunas interferencias. Además, se aprecia un ligero aumento de la entropía unos pocos segundos antes del inicio de la crisis, por lo que el número de permutaciones puede detectar con éxito el estado de la pre-crisis (pre-ictal) con una anticipación de unos pocos segundos. Por otra parte, se puede ver que después del final convulsiones (estado post-ictal), la regularidad de la señal de EEG difiere considerablemente de la anterior a la aparición de la convulsión, mostrando que el paciente no se recupera inmediatamente al estado normal.

2.2. LAS TABLAS DE CONTINGENCIA Y MEDIDAS DE ASOCIACIÓN ENTRE VARIABLES

Los fenómenos aleatorios se pueden representar mediante dos tipos de variables estadísticas:

- *Cuantitativas*: aquellas cuyos posibles valores son numéricos, por ejemplo, la estatura, el peso, la edad, la temperatura, etc. Estas se clasifican a su vez en variables continuas y discretas. Las variables continuas son aquellas que pueden tomar un conjunto infinito no numerable de valores y las discretas un conjunto finito o infinito numerable. Sin embargo, en la práctica la distinción se hace entre variables que toman muchos valores o relativamente pocos porque debido a la limitación de los instrumentos de medida todas las variables se miden de forma discreta.

- *Cualitativa*: (o categóricas o atributos o factores): aquellas cuyos valores son un conjunto de cualidades no numéricas a las que se le suele llamar categorías o modalidades o niveles, por ejemplo, el sexo (mujer, hombre), el color del pelo (moreno, rubio, castaño, pelirrojo), filosofía política (liberal, moderada, conservadora), estado civil (soltero, casado, divorciado, viudo), nivel de estudios (ninguno, primario, medio, universitario), etc.

Una propiedad deseable de las categorías es que sean exhaustivas (proporcionan suficientes valores para clasificar a toda la población) y mutuamente excluyentes (cada individuo se clasifica en una y solo una categoría).

Observemos que, a primera vista, la exhaustividad puede parecer muy restrictiva, por ejemplo, puede que deseemos saber qué opinan los liberales y conservadores frente a la legalización del aborto. En este caso el problema se resuelve redefiniendo la población mediante eliminación de los moderados.

Las variables categóricas podemos clasificarlas según el número de categorías:



- Dicotómicas: solo dos modalidades, por ejemplo, padecer una enfermedad (Si, No), Sexo (Hombre, Mujer), Resultado de una oposición (Aprobar, Suspender), en general los fenómenos de respuesta binaria, etc.
- Politémicas: más de dos categorías, por ejemplo, los fenómenos de respuesta múltiple, lugar de nacimiento, clase social, etc.

Consideremos una muestra de tamaño n clasificada según dos factores (variables cualitativas) A y B con I y J niveles, respectivamente. Representemos por n_{ij} ($i = 1, \dots, I; j = 1, \dots, J$) el número de individuos de la muestra que se clasifican simultáneamente en el nivel A_i de A y B_j de B. La tabla bidimensional siguiente, que contiene en cada una de sus $I \times J$ casillas las frecuencias observadas n_{ij} , se llama **tabla de contingencia** bidimensional y su nombre es debido a Pearson

		Factor B					n_i
		B_1	...	B_j	...	B_J	
Factor A	A_1	n_{11}	...	n_{1j}	...	n_{1J}	$n_{1.}$

	A_i	n_{i1}	...	n_{ij}	...	n_{iJ}	$n_{i.}$

	A_I	n_{I1}	...	n_{IJ}	...	n_{IJ}	$n_{I.}$
$n_{.j}$		$n_{.1}$...	$n_{.j}$...	$n_{.J}$	n

A partir de la tabla de contingencia se obtienen las distribuciones de frecuencias marginales y condicionadas de la siguiente forma:

Frecuencias marginales:

Factor A:

$$n_{i.} = \sum_{j=1}^J n_{ij} \quad (i = 1, \dots, I)$$

Factor B:

$$n_{.j} = \sum_{i=1}^I n_{ij} \quad (j = 1, \dots, J)$$

Verificando:

$$n = \sum_{i=1}^I \sum_{j=1}^J n_{ij} = \sum_{i=1}^I n_{i.} = \sum_{j=1}^J n_{.j}$$



Frecuencias condicionadas:

Factor A condicionado a $B = B_j$ ($j = 1, \dots, J$):

$$n_{i|j=n_{ij|j}} \quad i = 1, \dots, I$$

Factor B condicionado a $A = A_i$ ($i = 1, \dots, I$):

$$n_{j|i=n_{ij|i}} \quad j = 1, \dots, J$$

Observemos que para cada factor hay tantas distribuciones de frecuencias condicionadas como niveles tiene el otro factor, es decir, hemos definido en total $(I+J)$ distribuciones condicionadas de frecuencias observadas.

Recordemos que cualquier individuo de la población debe clasificarse en una y solo una de las $I \times J$ casillas de la tabla de contingencia. Si denotamos por p_{ij} a la probabilidad de que un individuo elegido aleatoriamente en la población se clasifique en el nivel A_i de A y en el nivel B_j de B, obtenemos la *función de probabilidad conjunta* de los factores A y B, verificando como cualquier función de probabilidad discreta que:

Para representarla, se usa una tabla de la siguiente forma, que tiene la misma estructura que la tabla de contingencia:

		Factor B					p_i
		B_1	\dots	B_j	\dots	B_J	
Factor A	A_1	p_{11}	\dots	p_{1j}	\dots	p_{1J}	$p_{1.}$
	\dots	\dots	\dots	\dots	\dots	\dots	\dots
	A_i	p_{i1}	\dots	p_{ij}	\dots	p_{iJ}	$p_{i.}$
	\dots	\dots	\dots	\dots	\dots	\dots	\dots
	A_I	p_{I1}	\dots	p_{Ij}	\dots	p_{IJ}	$p_{I.}$
$n_{.j}$		$p_{.1}$	\dots	$p_{.j}$	\dots	$p_{.J}$	p

$$\sum_{i=1}^I \sum_{j=1}^J p_{ij} = 1.$$

Al igual que para las frecuencias observadas, se definen las funciones de probabilidad marginales y condicionadas asociadas:

Funciones de probabilidad marginales:

Factor A:

$$p_{i.} = \sum_{j=1}^J p_{ij} \quad (i = 1, \dots, I)$$

Factor B:



$$p_{\cdot j} = \sum_{i=1}^I p_{ij} \quad (j = 1, \dots, J)$$

Verificando:

$$\sum_{i=1}^I p_{i\cdot} = \sum_{j=1}^J p_{\cdot j} = 1$$

Las distribuciones marginales dan información unidimensional sobre cada variable y no dicen nada sobre la asociación entre las dos variables.

Funciones de probabilidad condicionadas:

Factor A condicionado a B = B_j (j = 1, ... ,J):

$$p_{i|j} = p_{ij}/p_{\cdot j} \quad i = 1, \dots, I$$

Factor B condicionado a A = A_i (i = 1, ... ,I):

$$p_{j|i} = p_{ij}/p_{i\cdot} \quad j = 1, \dots, J$$

INDEPENDENCIA ENTRE VARIABLES.

Dos variables cualitativas son *estadísticamente independientes* si se verifica:

$$p_{ij} = p_i p_{\cdot j} \quad \forall i = 1, \dots, I; j = 1, \dots, J$$

Es decir, si y sólo si las probabilidades conjuntas son igual al producto de las probabilidades marginales.

De forma equivalente, dos variables son independientes si se verifica:

$$p_{j|i} = p_{\cdot j} \quad i = 1, \dots, I.$$

Esto significa que la función de probabilidad condicionada de B es igual que su función de probabilidad marginal. Es decir, dos variables son independientes cuando la probabilidad de clasificarse en la columna *j* es igual en todas las filas. Intuitivamente la independencia significa que los valores de una variable no vienen influidos por la modalidad que adopte la otra variable. De igual forma las variables son independientes si las distribuciones condicionales de A son iguales que su distribución marginal.

Ejemplo ilustrativo.

En el siguiente ejemplo suponemos que se quiere estudiar si existe relación entre fumar y padecer cáncer de pulmón, y que se conocen las probabilidades conjuntas asociadas a sus 4 combinaciones de categorías.



Consideraremos dos configuraciones distintas de probabilidades que son las que aparecen en las tablas siguientes.

Cáncer	Fumar		$p_{i.}$
	Si	No	
Si	0.3	0.3	0.6
No	0.2	0.2	0.4
$p_{.j}$	0.5	0.5	1

Tabla 2.1: Ejemplo de variables independientes

Cáncer	Fumar		$p_{i.}$
	Si	No	
Si	0.6	0.0	0.6
No	0.0	0.4	0.4
$p_{.j}$	0.6	0.4	1

Tabla 2.2: Ejemplo de variables dependientes

Se observa que en la primera tabla tener la enfermedad no está relacionado con fumar o no fumar (independencia), mientras que en la segunda tabla todos los fumadores tienen cáncer de pulmón y los no fumadores no lo tienen (dependencia estadística total).

En general, las probabilidades poblacionales vendrán estimadas a partir de datos muestrales. Así, para una muestra de tamaño n , si denotamos por n_{ij} al número de individuos de la muestra que se clasifican simultáneamente en el nivel A_i de A y en el nivel B_j de B , la función de probabilidad conjunta se estima mediante:

$$\hat{p}_{ij} = n_{ij}/n$$

y las marginales mediante:

$$\hat{p}_{i.} = n_{i.}/n$$

$$\hat{p}_{.j} = n_{.j}/n$$

ESTADÍSTICOS USUALES EN TABLAS DE CONTINGENCIA

Una de las principales utilidades de las tablas de contingencia es contrastar la presencia de asociación o dependencia entre variables cualitativas. Los estadísticos más usuales para el contraste de independencia mediante tablas de contingencia son el ji-cuadrado de Pearson (X^2) y el test de razón de verosimilitudes (también llamado G-test).

Contraste de independencia X^2 de Pearson: Consideremos una tabla de contingencia $I \times J$ generada por muestreo multinomial. Esto significa que la distribución conjunta del vector aleatorio que genera a las frecuencias observadas es una multinomial de parámetros las probabilidades poblacionales.



Supongamos que queremos contrastar la hipótesis nula de independencia:

$$H_0: p_{ij} = p_i p_{.j} \quad \forall i, j$$

Sean $\hat{m}_{ij} = n_i n_{.j} / n$ los estimadores de máxima verosimilitud de las frecuencias esperadas bajo la hipótesis de independencia. Para llevar a cabo este contraste, Pearson propuso el siguiente estadístico:

$$X^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}},$$

que bajo la hipótesis nula de independencia, tiene distribución de probabilidad asintótica (para tamaños muestrales grandes) ji-cuadrado con $(I-1)(J-1)$ grados de libertad, $X^2_{(I-1)(J-1)}$.

Por tanto, se rechaza la hipótesis de independencia al nivel α cuando se verifica

$$X^2_{obs} \geq X^2_{(I-1)(J-1); 1-\alpha}$$

Donde X^2_{obs} , denota el valor del estadístico del contraste para la muestra observada.

El test de razón de verosimilitudes para contrastar la hipótesis de independencia en una tabla de contingencia $I \times J$, se realiza mediante el estadístico:

$$G^2 = -2 \log A = 2 \sum_i \sum_j n_{ij} \log \frac{n_{ij}}{\hat{m}_{ij}}$$

siendo $\hat{m}_{ij} = n_i n_{.j} / n$ las frecuencias esperadas bajo la hipótesis de independencia.

Este estadístico, bajo la hipótesis de independencia, se distribuye asintóticamente como una v.a. X^2 con $(I-1)(J-1)$ grados de libertad. Por tanto, se rechaza la hipótesis de independencia al nivel α si se verifica:

$$G^2_{obs} \geq \chi^2_{(I-1)(J-1); 1-\alpha}$$

Una alternativa a los estadísticos tradicionales se presenta en el trabajo de Cressie y Read (1998). Se propone la siguiente familia de estadísticos dependiente de un parámetro λ :

$$CR(\lambda) = \frac{2}{\lambda(\lambda+1)} \sum_i \sum_j n_{ij} \left[\left(\frac{n_{ij}}{\hat{m}_{ij}} \right)^\lambda - 1 \right] \quad \text{con } 0 \leq \lambda \leq 1.$$

Dependiendo del valor de λ , este estadístico toma diferentes formas que incluyen a otros existentes. Si $\lambda = 0$ (límite cuando tiende a cero), se obtiene el estadístico G-test, y si $\lambda = 1$ se tiene el estadístico X^2 . Al igual que sucede con los estadísticos ji-cuadrado de Pearson y el G-test, bajo el supuesto de independencia, la familia de estadísticos propuesta por Cressie y Read (1998) se distribuye asintóticamente como una X^2 con $(I-$



1)(J-1) grados de libertad. Además recomiendan considerar $\lambda = 2/3$ para mejorar la adecuación de la aproximación asintótica.

Ejemplo ilustrativo.

Se quiere estudiar si existe asociación entre las variables “Fumar” y “Padecer cáncer de pulmón”. Para una muestra de 100 individuos se obtuvo la siguiente tabla de contingencia:

Cáncer	Fumar		$n_{i.}$
	Si	No	
Si	36	19	55
No	24	21	45
$n_{.j}$	60	40	n=100

Tabla 2.3: Tabla frecuencias observadas

Para calcular los estadísticos se necesitan conocer los valores de las frecuencias esperadas, las cuales toman un valor:

Cáncer	Fumar		$\hat{m}_{i.}$
	Si	No	
Si	33	22	55
No	27	18	45
$\hat{m}_{.j}$	60	40	n=100

Tabla 2.4: Tabla frecuencias esperadas

$$X^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}} = \frac{9}{33} + \frac{9}{22} + \frac{9}{27} + \frac{9}{18} = 1'5152$$

$$G^2 = 2 \sum_i \sum_j n_{ij} \log \frac{n_{ij}}{\hat{m}_{ij}} = 1'5145$$

$$CR(\lambda) = \frac{2}{\lambda(\lambda + 1)} \sum_i \sum_j n_{ij} \left[\left(\frac{n_{ij}}{\hat{m}_{ij}} \right)^\lambda - 1 \right] = 1'5146$$

Para decidir si las dos variables son dependientes o independientes los criterios de aceptación y rechazo de la independencia dependen del valor crítico de nuestro contraste al 95% de confianza, que viene dado por una chi-cuadrado con 1 grado de libertad y vale 7'879. Como se puede observar, en este caso no existe dependencia entre las variables, ya que los resultados de los estadísticos están muy por debajo de 7'879.

Ahora se plantea otro ejemplo con el mismo enunciado y el mismo tamaño muestral n=100, pero con distinta tabla de contingencia:

Cáncer	Fumar		$n_{i.}$
	Si	No	
Si	50	5	55
No	10	35	45
$n_{.j}$	60	40	n=100

Tabla 2.5: Tabla frecuencias observadas

Las frecuencias esperadas toman igual valores que el ejemplo anterior:

Cáncer	Fumar		$\hat{m}_{i.}$
	Si	No	
Si	33	22	55
No	27	18	45
$\hat{m}_{.j}$	60	40	n=100

Tabla 2.6: Tabla frecuencias esperadas

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}} = 48'65$$

$$G^2 = 2 \sum_i \sum_j n_{ij} \log \frac{n_{ij}}{\hat{m}_{ij}} = 48'5$$

$$CR(\lambda) = \frac{2}{\lambda(\lambda+1)} \sum_i \sum_j n_{ij} \left[\left(\frac{n_{ij}}{\hat{m}_{ij}} \right)^\lambda - 1 \right] = 53'41$$

La forma de aceptar o rechazar la independencia es igual que en el ejemplo anterior. Se ve que las variables presentan una gran dependencia, ya que los estadísticos están muy por encima de 7'879.

LIMITACIONES DE LA APROXIMACIÓN ASINTÓTICA.

Aunque los tres estadísticos anteriores (ji-cuadrado, G-test y $CR(\lambda)$) suelen ser adecuados para contrastar la independencia, conviene destacar algunas limitaciones:

- Según Cochran (1954), para que se pueda considerar correcta la significación calculada por el estadístico ji-cuadrado de Pearson, se debe cumplir que las frecuencias esperadas no sean muy pequeñas (inferiores a 5) más que en unas pocas casillas. Si es en muchas las casillas donde esto ocurre (más del 20%) se debe usar una prueba que no incluya aproximaciones, como la prueba exacta de Fischer.



En nuestro caso, esto significa que el tamaño de la ventana móvil " w " no debe ser excesivamente pequeño, puesto que se tendrían muchas frecuencias esperadas pequeñas.

- El test ji-cuadrado de Pearson es sensible al tamaño muestral, de manera que si las muestras son muy grandes, la prueba de independencia del test ji-cuadrado dará resultados significativos incluso donde, posiblemente, consideremos que las diferencias no sean en realidad significativas.

Esto se traduce en nuestro caso en que el tamaño de la ventana móvil " w " no debe ser tampoco excesivamente grande porque siempre concluiríamos dependencia entre las series en estudio.

- Si las variables poseen muchos niveles posiblemente la prueba no resulte de mucho interés, ya que es lógico esperar que se encuentren diferencias.

En nuestro caso, si la dimensión embedding seleccionada m es muy grande, tendremos una tabla de contingencia de $m!_{filas} \times m!_{columnas}$, lo cual no sería adecuado. Por tanto, usaremos como tamaño embedding $m=3$ o bien $m=4$.

- Koehler (1986) demuestra que, para el G-test, la aproximación asintótica a una ji-cuadrado es pobre cuando n/N es menor de 5, donde n es el tamaño de la muestra y $N = (I \cdot J)$ el número de celdas de la tabla de contingencia.

Esta limitación se traduce en que la selección del tamaño de la ventana móvil " w " y de la dimensión embedding " m " deberán mantener una determinada relación.

- En el manual StatXact (1991) se muestra que hay muchas configuraciones de datos para los que la aproximación asintótica resulta inadecuada. Entre ellas, pequeños conjuntos de datos, conjuntos de datos grandes pero no balanceados y conjuntos de datos poco densos o "sparse". En estos casos resulta necesario recurrir a test exactos o al menos a la estimación del p-valor exacto del contraste, como desarrollamos a continuación.

En el contexto de las series temporales y con el fin de evitar estas configuraciones "defectuosas", conviene trabajar además con series que presenten un alto grado de aleatoriedad, puesto que las deterministas darán lugar a tablas de contingencia no balanceadas.

Veamos un ejemplo ilustrativo de esta última limitación para la aproximación asintótica, así como las formas de subsanarla (véase StatXact (1991)). Para ello, previamente definiremos los conceptos de enumeración de Montecarlo y enumeración completa.

Hay tres tipos básicos de algoritmos para el cálculo de p-valores de un test, por ejemplo de independencia: la enumeración completa, la enumeración de Monte Carlo y p-valores asintóticos.

Los algoritmos de enumeración completa enumeran todos y cada uno de los resultados en el conjunto de referencia. Así que siempre produce el p-valor exacto. Su resultado es 100% exacto. Sin embargo, no garantiza poder resolver todos los problemas.



Algunos conjuntos de datos pueden ser demasiado grandes para la enumeración. Por lo cual también se usa el algoritmo de enumeración de Monte Carlo. Estos algoritmos enumeran un subconjunto aleatorio de todos los resultados posibles. Los algoritmos de Monte Carlo proporcionan una estimación del p-valor exacto, llamado el p-valor Monte Carlo, que puede ser tan exacto según sea necesario para el problema presente. Típicamente se proporciona una respuesta al 99% de exactitud, pero el usuario es libre de establecer el nivel de precisión. Para un nivel dado de precisión, el rango dentro del cual se encuentra el p-valor puede reducirse simplemente muestreando más los resultados de la serie de referencia. Por otra parte los métodos de Monte Carlo garantizan resolver cualquier problema, no importa el tamaño del conjunto de datos. Por tanto, proporcionan un respaldo robusto, fiable para las situaciones en las que los algoritmos de enumeración completa fallan. El p-valor Monte Carlo es, para todos los efectos prácticos, tan bueno como el p-valor exacto. Finalmente el p-valor asintótico está siempre disponible de forma predeterminada. Este cálculo es sencillo, y no se requiere algoritmos especiales.

En el siguiente ejemplo se muestra el espesor pared ventricular izquierda, medido por ecocardiografía, en 947 atletas que participan en 25 diferentes deportes, en Italia. Hubo 16 atletas con espesor de pared ≥ 13 mm, que es un indicativo de cardiomiopatía hipertrófica. Se desea determinar si existe alguna relación entre la presencia de esta condición y el tipo de deporte.

Una prueba apropiada para este conjunto de datos es la prueba de razón de verosimilitud (G-test). El valor de este estadístico es 32,495. El p-valor asintótico, en base a la prueba de razón de verosimilitud, es por lo tanto, la cola o área a la derecha de 32,495 a partir de una distribución X^2 con 24 grados de libertad. Este p-valor es 0,1152, **lo que indicaría que no existe una dependencia significativa entre ambas características.** Pero hay que fijarse que esta tabla no está balanceada y es sparse. Esto sugiere que no debemos confiar en el p-valor asintótico.

Deporte	Espesor		Total	Deporte	Espesor		Total
	≥ 13	< 13			≥ 13	< 13	
Levantamiento pesas	1	6	7	Buceo	1	10	11
Campo peso. eventos	0	9	9	Boxeo	0	14	14
Wrestling / Judo	0	16	16	Ciclismo	1	63	64
Tae kwon do	1	16	17	Waterpolo	0	21	21
Hockey sobre ruedas	1	22	23	Vela	0	24	24
Balonmano	1	25	26	Piragüismo	3	57	60
Cross-coum.esq	1	30	31	Esgrima	1	41	42
Esquí Esquí alpino	0	32	32	Tenis	0	47	47
Pentatlón	0	50	50	Remo	4	91	95
Patinaje sobre ruedas	0	58	58	Natación	0	54	54
Hípica	0	28	28	Fútbol	0	62	62
Bobsleigh	1	15	16	Track	0	89	89
Voleibol	0	51	51				

Tabla 2.7: Tabla frecuencias observadas



La estimación del p-valor usando Monte Carlo se hace del siguiente modo. Si M tablas son la muestra, y Q de ellas son al menos tan extremas como la tabla observada (en el sentido de tener una relación de probabilidad estadística mayor o igual a 32.495), la estimación del p-valor de Monte Carlo es:

$$\hat{p} = Q/M$$

La varianza de esta estimación se obtiene de forma sencilla por una Binomial:

$$var(\hat{p}) = \frac{p(1-p)}{M}$$

Para los datos de la Tabla anterior, hay una estimación de Monte Carlo de 0,0450 para el p-valor exacto, en base a 10.000 muestras aleatorias de la serie de referencia con un intervalo de confianza del 99% para el p-valor exacto de (0.0397, 0.0503), lo que indicaría que **en realidad sí existe una dependencia significativa entre las características en estudio.**

Aunque se puede demostrar matemáticamente que el p-valor exacto va a converger al correspondiente p-valor asintótico, cuando el tamaño de la muestra se hace infinitamente grande, esta propiedad no garantiza la exactitud del p-valor asintótico para todos los datos específicos establecidos. Hay muchas configuraciones diferentes de datos donde los métodos asintóticos realizan un cálculo pobre. Estos incluyen pequeños conjuntos de datos, grandes conjuntos de datos pero no balanceados y conjuntos de datos sparse (poco densos).

Debe quedar claro que es muy difícil predecir a priori si un determinado conjunto de datos es suficientemente grande como para realizar una aproximación asintótica. La definición de lo que constituye una gran muestra depende de la estructura de los datos y no puede ser caracterizado por una sola medida. Se ha observado incluso que una tabla de contingencia creada a partir de varios miles de observaciones, sin embargo puede producir p-valores asintóticos inexactos si posee muchas celdas con tamaños pequeños.

Aunque hay algunas ventajas de utilizar el método de Monte Carlo en lugar de utilizar el p-valor asintótico, indicar que en este proyecto se han seleccionado los parámetros del análisis con el fin de evitar todo lo posible esas configuraciones “defectuosas” y por tanto la necesidad de métodos exactos.

MEDIDAS DE ASOCIACIÓN ENTRE VARIABLES

En general, no sólo interesa estudiar si existe o no dependencia entre dos variables, sino que se desea conocer el grado de asociación o dependencia que presentan dichas variables.



El estadístico Ji-cuadrado de Pearson permite contrastar la independencia en una tabla de contingencia, pero no informa acerca del grado de asociación en el caso de que exista. Esto se debe a que su valor es muy sensible al tamaño muestral, de manera que para muestras muy grandes, aunque las frecuencias observadas difieran poco de las esperadas, el estadístico Ji-cuadrado tomará un valor alto y conducirá al rechazo de la independencia.

Entre las limitaciones del estadístico Ji-cuadrado, en el proyecto se ha destacado que es “sensible” al tamaño muestral. Para ilustrarlo, se propone el siguiente ejemplo:

Cosidérese la siguiente tabla de contingencia para una muestra de tamaño $n=10000$. Vamos a considerar una configuración idéntica a la del ejemplo propuesto en la página 31 en términos de probabilidades, pero aumentando el tamaño de la muestra.

Cáncer	Fumar		$n_{i.}$
	Si	No	
Si	3600	1900	5500
No	2400	2100	4500
$n_{.j}$	6000	4000	$n=10000$

Tabla 2.8: Tabla frecuencias observadas

La tabla de frecuencias esperadas, quedaría de la siguiente forma:

Cáncer	Fumar		$\hat{m}_{i.}$
	Si	No	
Si	3300	2200	5500
No	2700	1800	4500
$\hat{m}_{.j}$	6000	4000	$n=10000$

Tabla 2.9: Tabla frecuencias esperadas

$$X^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}} = 151'5152$$

$$G^2 = 2 \sum_i \sum_j n_{ij} \log \frac{n_{ij}}{\hat{m}_{ij}} = 151'4471$$

$$CR(\lambda) = \frac{2}{\lambda(\lambda+1)} \sum_i \sum_j n_{ij} \left[\left(\frac{n_{ij}}{\hat{m}_{ij}} \right)^\lambda - 1 \right] = 151'46346$$

Las variables serán dependientes si el valor de estos estadísticos es mayor que 7'879. Se puede ver que para $n=10000$ las variables son dependientes cuando para $n=100$ eran independientes.



Con el fin de minimizar el efecto del tamaño muestral, así como de tener medidas acotadas entre cero y uno, introducimos los siguientes coeficientes basados en el estadístico ji-cuadrado.

- Coeficiente de Contingencia:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Toma valores entre 0 y 1, pero difícilmente llega a 1. Su valor máximo depende del número de filas y columnas de la tabla de contingencia. Si toma valor cero indica independencia y cuando toma el valor máximo indica dependencia perfecta.

- Medida Phi cuadrado de Pearson:

$$\Phi = \sqrt{\frac{\chi^2}{n}}$$

En tablas de contingencia 2x2, toma valores entre 0 y 1, pero en el caso de tener más niveles puede tomar valores mayores que 1.

- V de Cramer:

$$V = \sqrt{\frac{\chi^2}{n \min\{(I-1)(J-1)\}}}$$

donde n es el total de observaciones, I el número de filas de la tabla de contingencia y J el número de columnas.

La V de Cramer es también una extensión del coeficiente Phi pero en este caso, y a diferencia del Coeficiente de Contingencia, si se encuentra normalizada. La V de Cramer oscila entre 0 y 1 (valores cercanos a 0 indican no asociación y los próximos a 1 fuerte asociación).

Una interpretación interesante de esta medida se puede encontrar en Wallis (2012), donde indica que el coeficiente V de Cramer mide el grado de transición entre una configuración equilibrada y la Identidad. O equivalentemente, que representa el flujo de información de la variable B hacia A . Si la información sobre B es irrelevante en la determinación de A , el coeficiente vale cero.



$$\Phi=0$$

F	a	-a
b	$1/2$	$1/2$
-b	$1/2$	$1/2$

$$\Phi=p$$

Φ	a	-a
b	$(p+1)/2$	$(1-p)/2$
-b	$(1-p)/2$	$(p+1)/2$

$$\Phi=1$$

I	a	-a
b	1	0
-b	0	1

Teniendo en cuenta la discusión anterior, hemos considerado preferible el uso de la V de Cramer frente al resto de coeficientes basados en la ji-cuadrado.

Para el caso de variables cuantitativas continuas, las medidas más conocidas de asociación o dependencia lineal son el coeficiente de correlación de Pearson y el de Spearman.

El coeficiente de correlación de Pearson es, quizá, el coeficiente más utilizado para estudiar el grado de relación lineal existente entre dos variables continuas. Se suele representar por r y se obtiene dividiendo la covarianza entre el producto de desviaciones típicas:

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

El coeficiente de correlación de Pearson toma valores entre -1 y 1 y es una medida simétrica: la correlación entre X e Y es la misma siempre que entre Y y X . Entre sus propiedades destacan:

- Carece de unidades de medida (adimensional).
- Es invariante para transformaciones lineales (cambio de origen y escala) de las variables.
- Sólo toma valores comprendidos entre -1 y 1.
- Si $r = 1$, existe una correlación positiva perfecta. El índice indica una dependencia total entre las dos variables denominada relación directa: cuando una de ellas aumenta, la otra también lo hace en proporción constante.
- Si $0 < r < 1$, existe una correlación positiva.
- Si $r = 0$, no existe relación lineal. Pero esto no necesariamente implica que las variables son independientes: pueden existir todavía relaciones no lineales entre las dos variables.
- Si $-1 < r < 0$, existe una correlación negativa.
- Si $r = -1$, existe una correlación negativa perfecta. El índice indica una dependencia total entre las dos variables llamada relación inversa: cuando una de ellas aumenta, la otra disminuye en proporción constante.

Es importante señalar que un coeficiente de correlación alto no implica causalidad. Dos variables pueden estar linealmente relacionadas (incluso muy relacionadas) sin que una sea causa de la otra. Por otra parte, adviértase que decimos "variables relacionadas linealmente". Esto significa que puede haber variables fuertemente relacionadas, pero no de forma lineal, en cuyo caso no proceder a aplicarse la correlación de Pearson.



Por ejemplo, la relación entre la ansiedad y el rendimiento tiene forma de U invertida; igualmente, si relacionamos población y tiempo la relación será de forma exponencial. En estos casos (y en otros muchos) no es conveniente utilizar la correlación de Pearson.

El coeficiente de correlación ρ de Spearman es el coeficiente de correlación de Pearson, pero aplicado después de transformar los datos originales en rangos. Toma valores entre -1 y 1, y se interpreta exactamente igual que el coeficiente de correlación de Pearson.

El coeficiente de Spearman suele utilizarse cuando alguna de las variables estudiadas es ordinal y/o se incumple el supuesto de normalidad.

Para finalizar esta sección, se va a introducir varias medidas de asociación basadas en el concepto de información mutua.

En teoría de la probabilidad, y en teoría de la información, la información mutua de dos variables aleatorias X e Y mide la dependencia de las dos variables, es decir, mide la reducción de la incertidumbre (entropía) de una variable aleatoria, X, debido al conocimiento del valor de otra variable aleatoria Y. Es por tanto una medida de dependencia o asociación, que denotaremos por $I(X;Y)$.

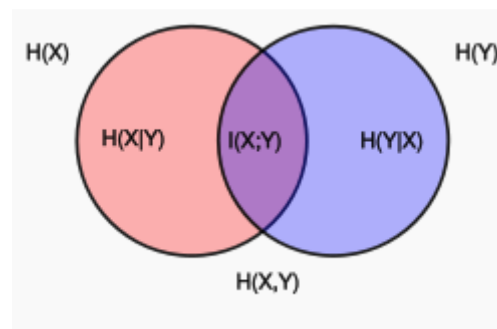
Dadas X e Y dos variables discretas, el coeficiente de información mutua entre ambos se define mediante:

$$I(X;Y) = \sum_{i=1}^n \sum_{j=1}^m P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)}$$

donde:

- $P(x_i, y_j)$: función puntual de probabilidad conjunta
- $P(x_i)$ y $P(y_j)$: marginales.

El siguiente diagrama muestra gráficamente el concepto de información mutua y su relación con la entropía:





donde:

- $H(X)$: Entropía de X = “Información que contiene X ”
- $H(Y)$: Entropía de Y = “Información que contiene Y ”
- $I(X;Y)$: Información que comparten X e Y .
- $H(X;Y)$: Entropía de (X,Y) = “Información que contiene X e Y conjuntamente”
- $H(X/Y)$: Entropía de X condicionada al conocimiento de Y .
- $H(Y/X)$: Entropía de Y condicionada al conocimiento de X .

Partiendo del diagrama anterior, el coeficiente de información mutua se puede obtener a partir del concepto de entropía, mediante la siguiente relación:

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

O equivalentemente:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Donde:

$$H(X) = - \sum_i p(x_i) \log p(x_i)$$

$$H(X,Y) = - \sum_{x,y} p(x,y) \log p(x,y)$$

$$H(X|Y) = - \sum_y p(y) \sum_x p(x|y) \log p(x|y)$$

El coeficiente de información mutua es una medida de dependencia puesto que $I(X,Y)=0$ si y solo si X e Y son independientes. Además, se trata de una medida simétrica y no negativa, pero su principal limitación es que no tiene cota superior fija.

Algunas variantes normalizadas del coeficiente de información mutua son las siguientes, que permiten acotar superiormente sus valores:

$$C_{XY} = \frac{I(X;Y)}{H(Y)} ; C_{YX} = \frac{I(X;Y)}{H(X)}$$

Los dos coeficientes anteriores no son necesariamente iguales.

En algunos casos, se desea obtener una medida simétrica, como la siguiente medida denominada Redundancia:

$$R = \frac{I(X;Y)}{H(X) + H(Y)}$$



que alcanza un valor mínimo de cero cuando las variables son independientes y un valor máximo cuando una de las variables es completamente redundante con el conocimiento de la otra:

$$R_{max} = \frac{\min(H(X), H(Y))}{H(X) + H(Y)}$$

Otra medida simétrica de especial utilidad es el Coeficiente de Incertidumbre (véase Witten y Frank 2005), dada por:

$$U(X, Y) = 2R = 2 \frac{I(X; Y)}{H(X) + H(Y)}$$

que representa una media ponderada de los dos coeficientes introducidos inicialmente.

Otras versiones normalizadas se proporcionan mediante las siguientes expresiones (Yao 2003, Strehl y Ghosh 2002).

$$\frac{I(X; Y)}{\min[H(X), H(Y)]'} \quad \frac{I(X; Y)}{H(X, Y)'} \quad \frac{I(X; Y)}{\sqrt{H(X)H(Y)}}$$

Por otra parte, muchas aplicaciones requieren el uso de una métrica, es decir, una medida de la distancia entre los puntos. Basados en el concepto de información mutua, encontramos las dos medidas de distancia universal siguientes:

$$D(X, Y) = 1 - I(X; Y)/H(X, Y)$$

y:

$$D'(X, Y) = 1 - \frac{I(X; Y)}{\max(H(X), H(Y))}$$

Se trata de métricas porque cumplen las propiedades de no-negatividad, simetría y desigualdad triangular. Además, son métricas universales en el sentido de que, si cualquier otra medida de distancia sitúa a las variables X e Y como “cercanas”, entonces también lo harán las distancias universales.

2.3. MEDIDAS DE “ASOCIACIÓN” Y “DISTANCIAS” ENTRE SERIES USANDO PERMUTACIONES

En esta sección introducimos algunas medidas que sirven para cuantificar el grado de asociación o dependencia existente entre dos series temporales. Nuestra idea se basa en el trabajo de Cánovas, Guillamón y Ruiz (2011), quienes desarrollan un test para contrastar la independencia entre dos series de tiempo haciendo uso de las permutaciones. Una vez codificadas las dos series en estudio, la información se recoge en una tabla de contingencia y el test de independencia se lleva a cabo mediante los



estadísticos usuales ji-cuadrado de Pearson o la razón de verosimilitud. Una de las principales ventajas de este método frente a otros propuestos anteriormente es su capacidad de detectar tanto relaciones lineales como no lineales.

En el contexto de los mercados eléctricos, por ejemplo, además de estudiar si existe o no asociación entre dos series de precios, resulta de interés estudiar si el grado de asociación que presentan los precios es el mismo a lo largo de toda la serie. O incluso si la asociación entre dos mercados eléctricos concretos es más fuerte que la de otro par de mercados.

Con este objetivo, en este proyecto proponemos realizar el estudio de la asociación o dependencia por ventanas móviles (en lugar de toda la serie), que podrían corresponderse con un periodo temporal (un mes, una estación, un año, etc.). Podría ocurrir, por ejemplo, que la asociación fuera mayor en los veranos y más débil en los otoños. Y por otra parte, teniendo en cuenta la discusión realizada en la sección anterior sobre medidas de asociación y de distancia, también proponemos comparar el grado de asociación entre pares de series a través de medidas normalizadas y no sensibles al tamaño muestral.

Recordemos de forma breve cómo se realiza la codificación de las series usando permutaciones y cómo se construiría la correspondiente tabla de contingencia.

Consideremos las series $(x_n)_{n=1}^T$ y $(y_n)_{n=1}^T$ y la serie bidimensional $(w_n)_{n=1}^T$, con $w_n = (x_n, y_n)$. Para π_i y π_j en S_m , decimos que la ventana $w_m(r)$ es de tipo $n_{i,j} = \pi_i \times \pi_j$ si y solo si $x_m(r)$ es de π_i -tipo y $y_m(r)$ es de π_j -tipo.

Una vez codificadas las series, toda la información se recoge en la tabla de contingencia:

	π_1	...	$\pi_{m!}$
π_1	$O_{1,1}$...	$O_{1,m!}$
...
$\pi_{m!}$	$O_{m!,1}$...	$O_{m!,m!}$

Tabla 2.10: Tabla de contingencia

Donde O_{ij} representan las frecuencias de aparición (valores observados) de cada símbolo $n_{i,j} = \pi_i \times \pi_j$.

$$O_{i,j} = \# \{w_m(r), r = 1, 2, \dots, T - m + 1: w_m(r) \text{ es del tipo } n_{i,j}\}$$

$$O_i = \sum_{j=1}^{m!} O_{ij} = \# \{x_m(r), r = 1, 2, \dots, T - m + 1: x_m(r) \text{ es del tipo } \pi_i\}$$



$$O_{.j} = \sum_{i=1}^{m!} O_{ij} = \# \{y_m(r), r = 1, 2, \dots, T - m + 1 : y_m(r) \text{ es del tipo } \pi_j\}$$

Entonces, la probabilidad asociada a cada símbolo se obtiene mediante:

$$p(\eta_{i,j}) = \frac{\# \{w_m(r), r = 1, 2, \dots, T - m + 1 : w_m(r) \text{ es del tipo } n_{i,j}\}}{T - m + 1}$$

y las probabilidades marginales:

$$p_1(\pi_i) = \frac{\# \{x_m(r), r = 1, 2, \dots, T - m + 1 : x_m(r) \text{ es del tipo } \pi_i\}}{T - m + 1}$$

y

$$p_2(\pi_j) = \frac{\# \{y_m(r), r = 1, 2, \dots, T - m + 1 : y_m(r) \text{ es del tipo } \pi_j\}}{T - m + 1}$$

El estudio se basa en el hecho de que bajo la independencia, se da la siguiente igualdad:

$$p(\pi_i \times \pi_j) = p_1(\pi_i) \cdot p_2(\pi_j)$$

para todo i, j . Por lo tanto, se propone la siguiente prueba con los códigos de permutación descritos anteriormente:

$$\begin{cases} H_0: p(\pi_i \times \pi_j) = p_1(\pi_i) \cdot p_2(\pi_j) \\ H_1: \text{Cualquier otro caso} \end{cases}$$

Para entender mejor la forma de construir la tabla de contingencia y como se realiza el cálculo de las frecuencias esperadas y observadas, se va a proponer un ejemplo.

Ejemplo ilustrativo. Consideremos dos series de datos simulados, $X(n)$ e $Y(n)$, ambas de tamaño 10000 datos.

$$X(n) = \{2'307; 1'047; 1'891; 2'009; 1'943; 2'264; 2'299; 1'128; 0'852; 0'621, \dots\}$$

$$Y(n) = \{-5'054; -4'253; -2'569; -2'016; -1'523; 0'020; 1'274; 1'413; 2'044; 3'027, \dots\}$$

Supongamos que tenemos una dimensión de embedding $m=3$ y un tamaño de ventana móvil deslizante $w=1000$. Por tanto, tendríamos un total de 10 ventanas móviles de tamaño 1000 sin solapar a lo largo de la serie.



Al tener $m=3$, tenemos $3! = 6$ permutaciones distintas, que serían $\{1,2,3\}, \{1,3,2\}, \{2,1,3\}, \{2,3,1\}, \{3,1,2\}$ y $\{3,2,1\}$ y que denotaremos por $\pi_1, \pi_2, \dots, \pi_6$, respectivamente. La forma de proceder es la siguiente:

1. Seleccionamos los 1000 primeros datos de cada serie, lo que corresponde a la primera ventana móvil de tamaño 1000 de la serie.
2. Posteriormente codificamos los 3 primeros datos (subventana de tamaño embedding). En el caso de $X(n) = \{2'307; 1'047; 1'891\}$ obtenemos la permutación $\pi_5 = \{3,1,2\}$. En el caso de $Y(n)$ sería la permutación $\pi_1 = \{1,2,3\}$. Por tanto, en la celda de valores $O_{5,1}$ contaríamos una nueva aparición.
3. Repetimos el paso 2, $w - m + 1$ veces para la primera ventana deslizante. Es decir, en este caso ordenaríamos del 2º al 4º dato de $X(n)$ e $Y(n)$, obteniendo los símbolos $\{1,2,3\}$ para $X(n)$ y $\{1,2,3\}$ para $Y(n)$, con lo cual el valor observado de la celda $O_{1,1}$ aumentaría una unidad.

La tabla de contingencia para la primera ventana móvil sería:

	π_1	π_2	π_3	π_4	π_5	π_6	$O_{i.}$
π_1	59	32	30	34	57	28	240
π_2	27	16	20	16	38	13	130
π_3	27	22	18	20	41	15	143
π_4	31	12	18	11	25	18	115
π_5	55	32	26	36	53	40	242
π_6	40	20	17	15	23	13	128
$O_{.j}$	239	134	129	132	237	127	998

Tabla 2.11: Tabla contingencia Ejemplo

donde $O_{1,1} = 59$, ..., y $O_{6,6} = 13$.

Este proceso se repetiría para cada una de las 10 ventanas móviles de tamaño 1000 y sin solapar que aparecen a lo largo de la serie.

A partir de la tabla de contingencia, las frecuencias esperadas bajo la hipótesis de independencia se obtienen de la siguiente manera:

$$fe_{i,j} = \frac{(O_{i.} \times O_{.j})}{n}$$



Por ejemplo, la frecuencia esperada para la entrada (1,1) de la tabla de contingencia anterior sería:

$$f_{e1,1} = \frac{240 \times 239}{998} = 57'47$$

Inicialmente se han propuesto para este proyecto como medidas de asociación los siguientes estadísticos ya mencionados:

- Chi-cuadrado de Pearson

$$X^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$$

Donde n_{ij} son los valores observados y \hat{m}_{ij} los valores esperados

- El test de razón de verosimilitudes

$$G^2 = -2 \log \Lambda = 2 \sum_i \sum_j n_{ij} \log \frac{n_{ij}}{\hat{m}_{ij}}$$

- Cressie-Read

$$CR(\lambda) = \frac{2}{\lambda(\lambda + 1)} \sum_i \sum_j n_{ij} \left[\left(\frac{n_{ij}}{\hat{m}_{ij}} \right)^\lambda - 1 \right]$$

donde $0 \leq \lambda \leq 1$.

Recordar que estos estadísticos no tienen sus valores acotados en un intervalo fijo y que dependen de los parámetros:

- n: tamaño de la serie
- w: tamaño de la ventana móvil
- m: dimensión del embedding

Por ese motivo, sólo usaremos estas medidas cuando el objetivo sea contrastar la hipótesis de independencia entre series, puesto que disponemos de la distribución asintótica que permite establecer las regiones de aceptación y rechazo.

Sin embargo, cuando el objetivo sea medir la fuerza de la asociación y que sea comparable entre pares de series, independientemente de los parámetros embedding y ventana móvil seleccionados, usaremos las medidas de asociación y distancias normalizadas descritas en la sección anterior.

- Distancia basada en V-Cramer

$$D_{V-Cramer} = 1 - V$$

$$V = \sqrt{\frac{X^2}{n \min\{(I-1)(J-1)\}}}$$



donde n es el total de observaciones, I el número de filas de la tabla de contingencia y J el número de columna.

- Distancia basada en el Coeficiente de Incertidumbre:

$$D_U = 1 - U$$

$$U(X, Y) = 2R = 2 \frac{I(X; Y)}{H(X) + H(Y)}$$

$$I(X; Y) = H(X) + H(Y) - H(X, Y)$$

donde $H(X)$ y $H(Y)$ son las entropías de las series.

- Distancia Métrica Universal (1):

$$D(X, Y) = 1 - I(X; Y)/H(X, Y)$$

- Distancia Métrica Universal (2):

$$D'(X, Y) = 1 - \frac{I(X; Y)}{\max(H(X), H(Y))}$$

Destacar que estas medidas se usarán siempre que queramos cuantificar el grado de asociación entre series, no para contrastar la independencia, puesto que no disponemos de la distribución asintótica para establecer las regiones de aceptación y rechazo (salvo en el caso de la V de Cramer).



3. SOFTWARE UTILIZADO

Debido a la complejidad de los cálculos necesarios en este proyecto y a la ausencia de un software que cumpliera con los requisitos deseados, se ha tenido que diseñar una serie de programas para realizar los análisis requeridos. Principalmente se implementaron dos programas: Dependencias y Retardo, aunque también se diseñaron algunos programas auxiliares que ayudaron en algún punto del proyecto.

La variedad de paquetes de cómputo para el análisis de series temporales es grande (por ejemplo, SAS, SPSS, Eviews, RATS, Stata o Matlab). En este caso, interesa un software que pueda analizar una serie de datos de gran longitud (tanto los datos simulados como la de los precios de los distintos mercados son series de 100000 datos), y también que se adapte a nuestras necesidades. También se buscaban programas que fueran flexibles, porque algunos objetivos del proyecto cambiaban a medida que se iba avanzando, por lo cual era necesario modificar el software. Matlab es una herramienta, que al trabajar con vectores, permite trabajar con multitud de datos de una forma bastante rápida y eficiente. Además da la posibilidad de hacer un programa “a nuestra medida”, pudiendo cambiar parámetros, estadísticos, gráficos, etc., en cualquier momento. Por todo ello se eligió diseñar los programas en Matlab.

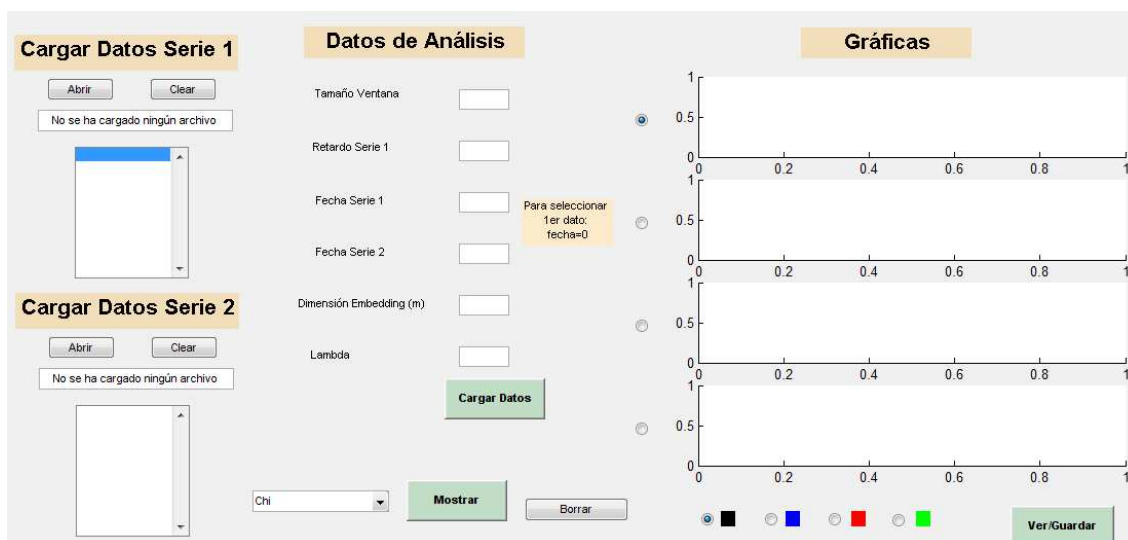
A continuación se va a explicar el funcionamiento de cada uno de los programas.

3.1. Manual de usuario del programa Dependencias

Este programa se encargará de analizar si dos series temporales (un par de series) presentan algún tipo de dependencia. Esta dependencia será estudiada con varios estadísticos. Se podrá elegir el tamaño de ventana (w), se podrá retrasar cualquier serie para analizar si hay dependencias con algún tipo de retardo, elegir la fecha de comienzo de las series, la dimensión del embedding (m) y el parámetro λ (necesario para el cálculo de Cressie-Read). También si se desea se puede hacer una representación gráfica de la dependencia del estadístico deseado y se pueden guardar los datos para representarlos posteriormente, sin tener que volver a realizar el estudio.



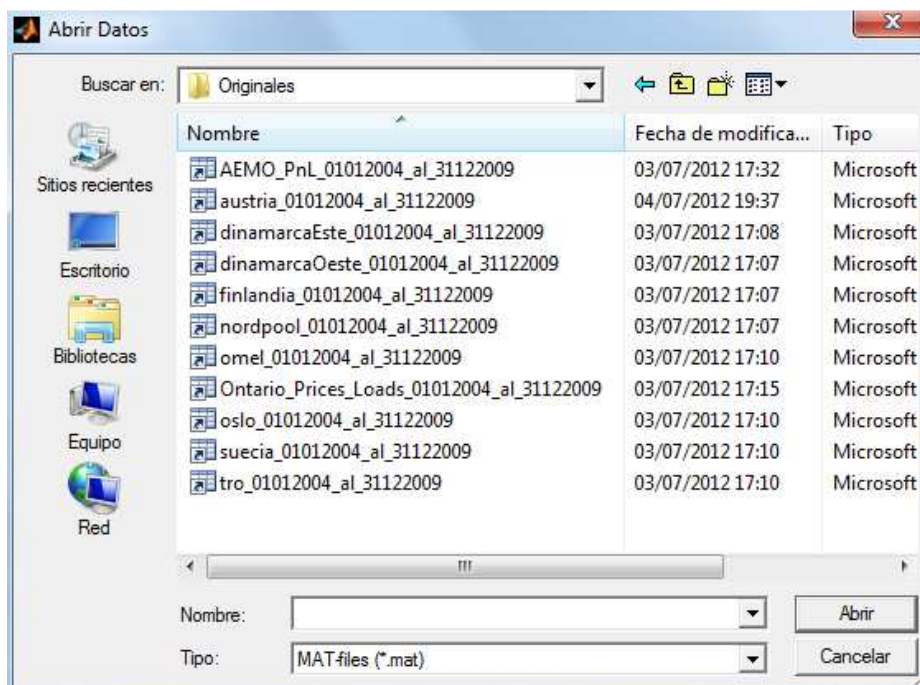
Al invocar al programa, aparecerá la siguiente interfaz gráfica:



Para entender mejor el funcionamiento del programa, se va a explicar cada una de las partes que lo forman por separado:

1. Cargar datos Serie 1/2

En la parte izquierda del programa será donde se seleccionen las series sobre las que se va a realizar el estudio de dependencia. Al pulsar en abrir, se abrirá la siguiente ventana:



Aquí se seleccionará el fichero que contiene la serie que deseamos estudiar. Los ficheros de matlab (.mat) pueden contener multitud de variables. Nuestras series son guardadas como variables en forma de vector columna con números de tipo double. Como ejemplo, se va a comparar el mercado OMEL con Nordpool.



- Tras seleccionar las dos series, se seleccionarán los parámetros necesarios para realizar el análisis. Estos parámetros se encuentran en la parte central de la interfaz, en la sección “Datos de Análisis”. Aquí se especificará: el tamaño de ventana (w), se podrá retrasar cualquier serie para analizar si hay dependencias con algún tipo de retardo (Retardo Serie 1), elegir la fecha de comienzo de ambas series (Fecha Serie 1 y 2), la dimensión del embedding (m) y el parámetro λ . Tras rellenar estos elementos, se pulsará el botón “Cargar Datos” y comenzará el análisis. Cuando concluyan los cálculos, aparecerá un mensaje “Se han cargado los datos” al lado del botón de “Cargar Datos”.

La interfaz muestra los siguientes campos de entrada:

- Tamaño Ventana: 2190
- Retardo Serie 1: 0
- Fecha Serie 1: 0
- Fecha Serie 2: 0
- Dimensión Embedding (m): 3
- Lambda: 0.666

Debajo de los campos, hay un mensaje "SE HAN CARGADO LOS DATOS" rodeado por un círculo rojo, y un botón "Cargar Datos". A la derecha de los campos de fecha, hay un texto amarillo que dice "Para seleccionar 1er dato: fecha=0".

- Tras esto, se pueden mostrar los resultados en una gráfica, mediante un desplegable y los botones situados en la parte central inferior de la interfaz. Se puede elegir entre los distintos estadísticos calculados a través del desplegable:

Se muestra un menú desplegable con las siguientes opciones:

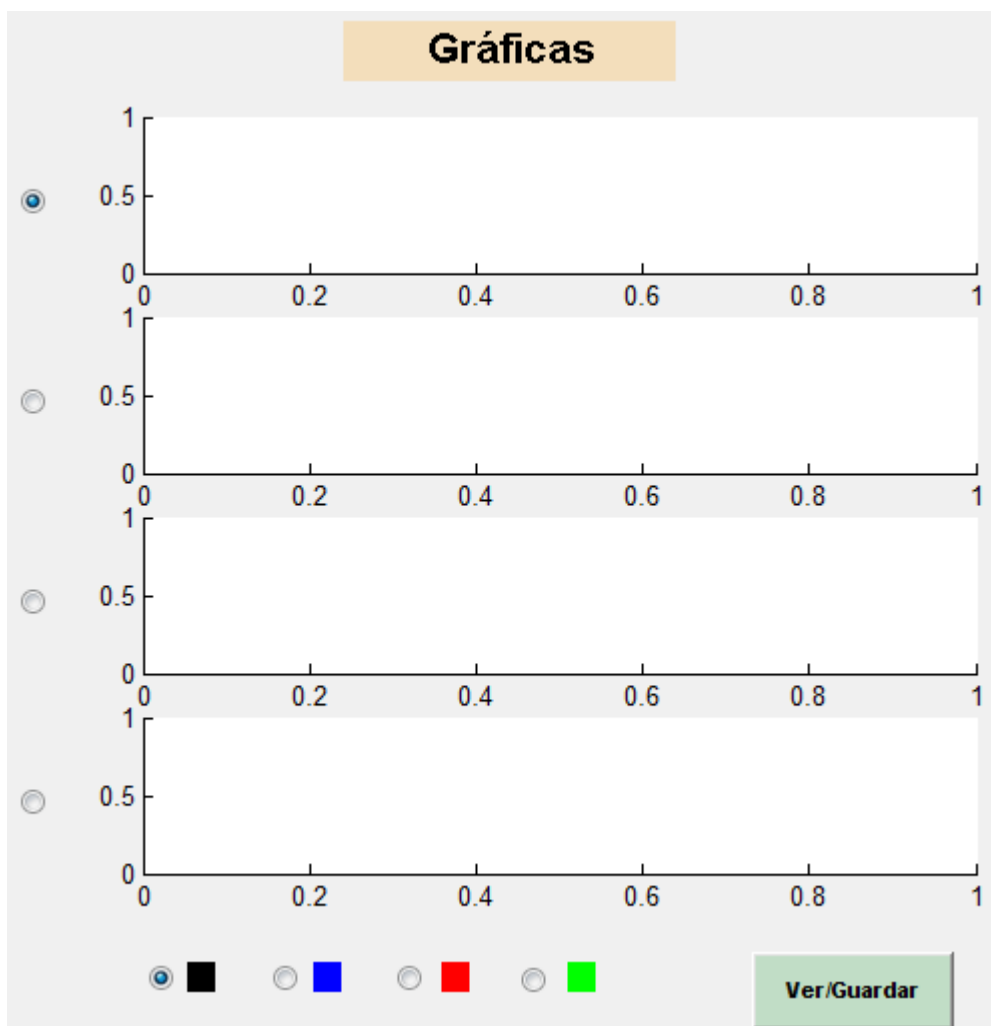
- Chi
- CR
- G-Test
- Spearman
- Pearson
- p_valor Spearman
- p_valor Pearson
- Dist V-Cramer
- Dist Coef Incertidumbre
- Dist Univ 1
- Dist Univ 2
- Serie 1 Original
- Serie 2 Original
- Chi

Debajo del menú, hay un botón "Mostrar" y un botón "Borrar".

Los estadísticos disponibles son, por este orden: Chi-cuadrado, Cressie-Read, G-test, correlación lineal de Spearman, correlación lineal de Pearson, p-valor de la correlación lineal de Spearman, p-valor de la correlación lineal de Pearson,

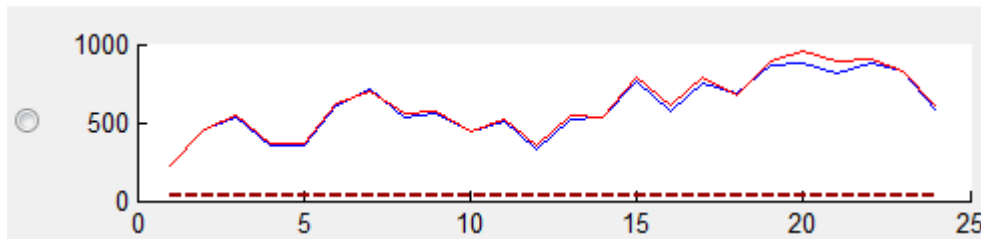
Distancia V de Cramer, Distancia del Coeficiente de Incertidumbre, Distancia Universal (1) y Distancia Universal (2). También es posible representar las series originales del estudio.

4. Después de seleccionar el estadístico, se pulsa el botón “Mostrar”, pero previamente debemos seleccionar la gráfica donde se desea representar los datos calculados. Para ello, se hará clic en el botón situado a la izquierda de cada gráfica. También se puede seleccionar el color para representar los resultados. Todo esto está situado en la parte derecha de la interfaz:



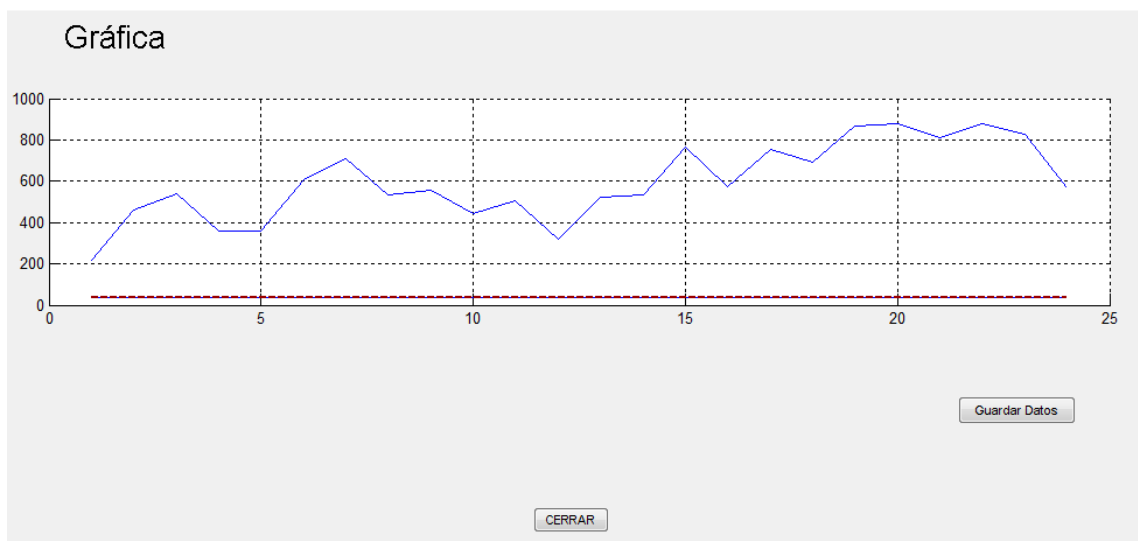
Se pueden mostrar todos los estadísticos que se deseen en una misma gráfica, aunque se recomienda que sean, como máximo, 4 estadísticos, de forma que se pueda representar cada estadístico con un color distinto. También es aconsejable no representar estadísticos de distinta naturaleza en la misma gráfica. Por ejemplo, los estadísticos Chi-Cuadrado, G-test y Cressie-Read tienen unos criterios de aceptación de dependencia distintos a los de correlación lineal. Para saber si las series son dependientes o independientes, en la gráfica se muestra una línea discontinua que nos indica las regiones de aceptación y rechazo de la hipótesis de independencia al 95% de confianza.

Por ejemplo, vamos a mostrar en una misma gráfica los estadísticos Chi-cuadrado (azul) y G-test (rojo):



Al estar por encima de la línea discontinua, vemos que ambos estadísticos muestran que existe una gran dependencia entre los mercados de España (OMEL) y países nórdicos (Nordpool). Destacar que estos datos no han sido desestacionalizados, razón por la cual todos estos análisis tendrán como resultado la dependencia de los mercados eléctricos. También mencionar, que cuando se ha querido analizar la dependencia mediante los métodos de correlación lineal, se ha utilizado los p-valores de ambos estadísticos, ya que con ello disponemos de una forma exacta para la aceptación o no de la dependencia. Si se desea eliminar la gráfica para representar otros datos, está el botón “Borrar” en la parte central, al lado del botón “Mostrar”.

5. Por último, en la parte inferior derecha está el botón “Ver/Guardar”. Al pulsarlo, se abrirá una segunda interfaz:

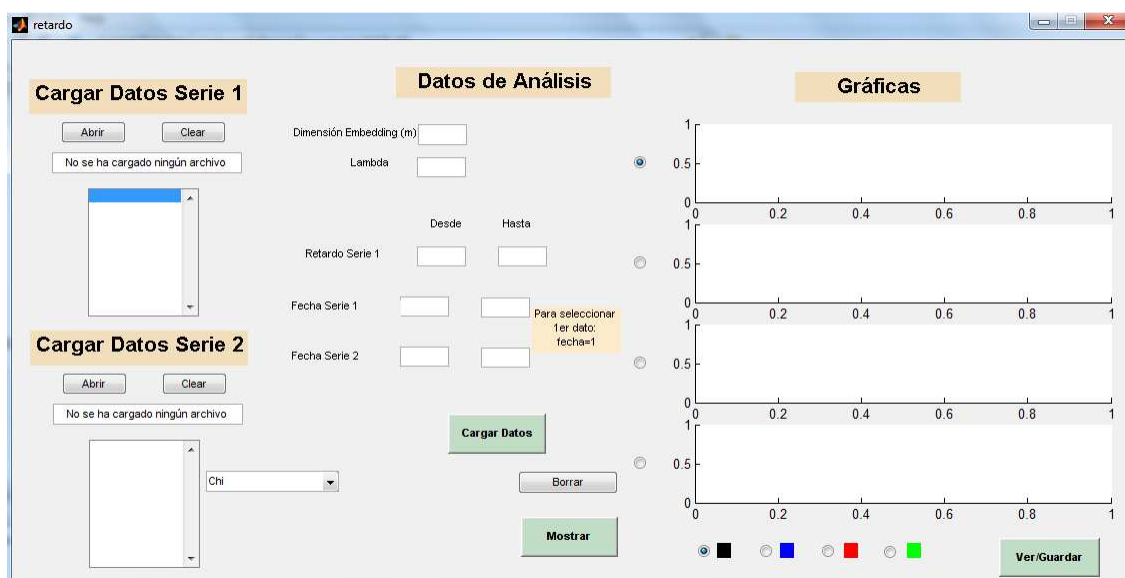


En ella se podrá ver de forma más clara la representación de los estadísticos. También se podrán guardar los resultados. En realidad, lo que se almacena es la tabla donde se han ido guardando todos los valores de los estadísticos. Si se guarda dicha tabla, se podrá volver a representar los resultados del análisis, mediante otro programa.

3.2. Manual de usuario del programa Retardo

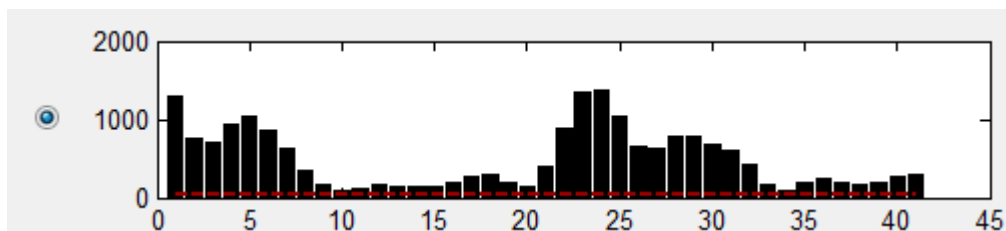
Este programa se encargará de identificar el retardo que proporciona mayor dependencia funcional entre dos series. La forma de cargar las series, de introducir los parámetros necesarios y de notificar que el cálculo ha finalizado es muy similar al del programa Dependencias, aunque en este caso se introducirá la fecha de comienzo y final de la serie. Las longitudes de ambas series deben coincidir, de forma que el programa avisará con un error y no dejará proseguir el cálculo en caso contrario.

Al invocar al programa, aparecerá la siguiente interfaz gráfica:



Otra de las diferencias respecto al programa Dependencias es que no hay que introducir el tamaño de ventana móvil, pues coincide con la longitud de la serie. La representación gráfica mostrará el valor del estadístico seleccionado para cada uno de los distintos retardos.

Al mostrar los resultados, el gráfico obtenido es distinto al del programa Dependencias. Por ejemplo, vamos a ver en qué momento los mercados de Nordpool y OMEL presentan más dependencia. Se ha elegido el año 2004, con una dimensión de embedding de 3 y se desea que muestre el intervalo de retardos desde 0 hasta 40:



Se puede apreciar que las mayores dependencias son para el retardo 0 y el retardo 24, es decir, sincronizadas en el tiempo y justo un día después.

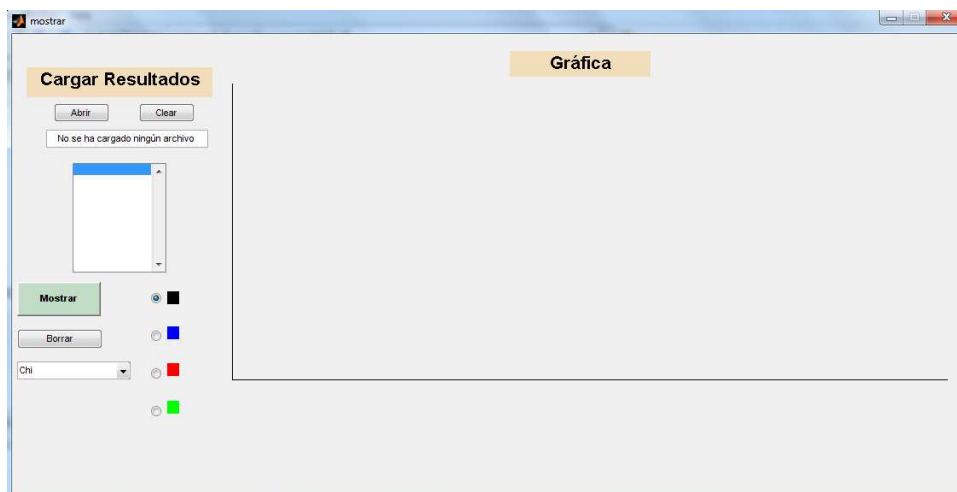


Hay que tener en cuenta al observar los resultados que la numeración del eje X está desplazado en una unidad. Esto se debe a que Matlab empieza a mostrar los resultados a partir de 1, a pesar de que se desee mostrar el retardo 0. Por ejemplo, en el gráfico superior, el primero de los resultados se corresponde para un retardo 0, aunque lo más intuitivo al mirar la gráfica sería pensar que el primer dato se corresponde a un retardo 1.

En este programa también está la opción de ver y guardar los datos del análisis realizado. Para ello solo se debe pulsar el botón "Ver/Guardar" y proceder de la misma manera que en el programa Dependencias.

3.3. Manual de usuario del programa Mostrar

El programa Mostrar es un programa de apoyo. Este programa nos permitirá mostrar los resultados de un estudio anterior, siempre y cuando hayamos guardado la variable "tabla". Este programa tiene la siguiente interfaz:



La forma de seleccionar la serie, el estadístico a representar y la selección del color es la misma a la utilizada en los programas Dependencias y Retardo.

3.4. Implementación de la metodología en Matlab

Tras ver cómo se utilizan cada uno de los programas diseñados necesarios para realizar el análisis de las series temporales, se explicará como se ha realizado el desarrollo de dichos programas.

3.4.1. Programa Dependencias

En realidad el programa Dependencias es solo una interfaz que se encarga de obtener los parámetros necesarios para llamar a otras funciones, pasándole como argumento los valores de dichos parámetros. Posteriormente, nos permite representar dichas funciones



en los “axes” de Matlab. Al pulsar el botón “Cargar Datos”, se obtienen los parámetros: tamaño de ventana (w), lambda, dimensión del embedding (m), las fechas de inicio de ambas series y el retardo de la serie número 1; y se almacenan en su variable respectiva de la siguiente forma:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
m = str2double(get(handles.edit12,'String'));
retserie = str2double(get(handles.edit5,'String'));
ventana = str2double(get(handles.edit4,'String'));
lambda = str2double(get(handles.edit13,'String'));
fecha1 = str2double(get(handles.edit6,'String'));
fecha2= str2double(get(handles.edit11,'String'));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Con estos datos, se invoca a la función “tablabidim”, pasándole como argumentos las variables anteriormente obtenidas. Destacar que esta función devuelve una tabla, donde están almacenados todos resultados de los estadísticos, medidas de asociación y distancias. Además, posee otros valores, como los límites de aceptación de dependencia, que son necesarios para su representación.

La invocación de la función “tablabidim” se tiene que hacer de la siguiente forma:

```
tablabidim2(serie1,serie2,m,ventana,retardo,landa,fecha1,fecha2)
```

Donde “serie1” y “serie2” son las series que vamos a analizar. Todos los demás parámetros son los ya mencionados. Este programa es el que se encarga de codificar los datos y completar la tabla de contingencia, para después calcular los estadísticos y distancias.

Para la codificación de los datos, el programa “tablabidim” llama a la función “ordenar”. Este programa devuelve los datos ordenados y necesita que le pasen como argumentos la serie a analizar y la dimensión del embedding.

```
permuta = ordenar(serie,m)
```

El programa “ordenar” se encarga basicamente de la codificación de los datos y lo hace con la función “sort”, que devuelve la serie ordenada.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function permuta = ordenar(serie,m)

tam=length(serie);
longitud=tam-m+1;

for t=1:longitud

    serieaux = serie(t:t+m-1);
    [a,b] = sort(serieaux);

    permuta(t,:)=b;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```



Con el código siguiente hacemos se calcula la frecuencias observadas de cada permutación (por ejemplo el {1,3,2}) y con la función “apilar” se coloca todo esto en forma de tabla.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for t=1:longitud
    for t1=1:factorialm
        if permutax(t,:)==tablax(t1,1:m)
            for t2=1:factorialm
                if permutay(t,:)==tablay(t2,1:m)
                    frec(t1,t2)=frec(t1,t2)+1;
                    break;
                end
            end
        end
        break;
    end
end
end
todo=apilar(frec,m);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

A continuación, se completan dos tablas de contigencia: una con la frecuencias observadas y otra con las frecuencias esperadas.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
indice2=0;
indice1=0;
for t1=1:factorialm
    if sum(frec(t1,:))~=0
        indice1=indice1+1;
        almacen1(indice1)=t1;
    end
    if sum(frec(:,t1))~=0
        indice2=indice2+1;
        almacen2(indice2)=t1;
    end
end
total_obs=sum(sum(frec));
for t1=1:indice1
    for t2=1:indice2
        tabla_new(t1,t2)=frec(almacen1(t1),almacen2(t2));
        frec_obs(t1,t2)=frec(almacen1(t1),almacen2(t2));
    end
end

for t1=1:indice1
    for t2=1:indice2

        frec_esperada(t1,t2)=sum(tabla_new(t1,:))*sum(tabla_new(:,t2))/total_obs;
    end
end
grad_lib=(indice1-1)*(indice2-1);

```



%%%

Y para finalizar, se calcula el valor de los estadísticos y de las distancias. Después dichos valores son introducidos en distintas filas de la variable “tabla”.

%%%

```

suma_chi=0;
cr=0;
gtest=0;
for t1=1:indice1
    for t2=1:indice2
        suma_chi=suma_chi+(frec_obs(t1,t2)-
            frec_esperada(t1,t2))^2/frec_esperada(t1,t2);
        cr=cr+frec_obs(t1,t2)*((frec_obs(t1,t2)/frec_esperada(t1,t2)
            )^landa-1);
        if frec_obs(t1,t2)~=0
            gtest=gtest+frec_obs(t1,t2)*log(frec_obs(t1,t2)/frec_espera
                da(t1,t2));
        end
    end
end

k=factorialm;

ncramer=ventana-m+1;

vcramer=sqrt(suma_chi/(ncramer*(k-1)));
dvcramer=1-vcramer;

hxy=0;
for q=1:factorialm
    for t=1:factorialm
        ni=ni+frec_obs(q,t);
        nj=nj+frec_obs(t,q);
        if frec_obs(t,q)~=0
            hxy=hxy-
                (frec_obs(t,q)/ncramer)*log(frec_obs(t,q)/ncramer);
        end
    end

    tabla(14,q)=ni;
    tabla(13,q)=nj;
    ni=0;
    nj=0;
end

hx=0;
hy=0;
for q=1:factorialm
    if tabla(14,q)~=0
        hx=hx-(tabla(14,q)/ncramer)*log(tabla(14,q)/ncramer);
    end
    if tabla(13,q)~=0
        hy=hy-(tabla(13,q)/ncramer)*log(tabla(13,q)/ncramer);
    end
end

```



```

end

ixy=hx+hy-hxy;
d1=1-(ixy/hxy);

if hx>=hy
    d2=1-(ixy/hx);
else
    d2=1-(ixy/hy);
end

dixy=1-(2*ixy/(hx+hy));

gtestfinal=gtest*2;
[RHO1,PVAL1]=corr(serielaux,serie2aux,'type','Spearman');
[RHO2,PVAL2]=corr(serielaux,serie2aux,'type','Pearson');

crfinal=cr*2/(landa*(landa+1));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

3.4.2. Programa Retardo

Al igual que ocurría con el programa Dependencias, el programa Retardo es solo una interfaz que se encarga de obtener los parámetros necesarios para llamar a otras funciones. Puede representar los distintos retardos para el estadístico deseado. Los parámetros que se obtienen al pulsar el botón “Cargar Datos” son: lambda, dimensión del embedding (m), las fechas de inicio y fin de ambas series y el rango de retardos que se desea mostrar. La forma de obtenerlos y almacenarlos es la misma que en el programa Dependencias.

La función “tablabidretardo” devuelve una tabla donde se almacenan los resultados de los estadísticos y distancias.

La invocación de la función “tablabidretardo” se hace de la siguiente forma:

```

tablabidretardo(puntos1,puntos2,m,retseriei,retserief,lambda,fechali,fechalf,fecha2i,fecha2f);

```

Donde “puntos1” y “puntos2” son las series que vamos a analizar. Este programa, al igual que “tablabidim”, codifica los datos y completa la tabla de contingencia, para después calcular los estadísticos y distancias.

El programa “tablabidretardo” llama a la función “ordenar” para codificar los datos. La función “ordenar” es la misma que la utilizada en el programa Dependencias.

La forma de rellenar la tabla de contingencia, de calcular las frecuencias esperadas, de calcular el valor de los estadísticos de las distancias y de introducir estos en distintas filas de la variable “tabla” se realiza del mismo modo en el que se realizaba en “tablabidim”.



4. FUNCIONAMIENTO DE LAS MEDIDAS PROPUESTAS PARA DATOS SIMULADOS

En la primera parte de este capítulo se propondrán pares de series, algunos dependientes y otros independientes para analizar la fiabilidad de nuestras medidas en la detección de asociaciones. En la segunda parte se analizará cómo afecta la presencia de estacionalidad en el estudio de la dependencia de series. En la última parte del capítulo se intentará identificar el retardo en el tiempo que proporciona mayor dependencia funcional entre las series. En todos los casos, las series en estudio se generan mediante simulación.

4.1. EFICACIA EN LA DETECCIÓN DE DEPENDENCIA FUNCIONALES

Se va a estudiar la eficacia de las medidas propuestas en la detección de dependencias tanto lineales como no lineales. Para ello calcularemos el porcentaje de fallos en la detección de asociaciones y compararemos los resultados obtenidos frente a utilizar las medidas de asociación lineales usuales (correlación de Pearson y Rho de Spearman).

Primero, se va proponen seis pares de modelos independientes, los cuales son:

$$1. X_1(n) = 0'9 \cdot X_1(n-1) + u(n)$$

$$Y_1(n) = 0'9 \cdot Y_1(n-1) + v(n)$$

$$2. X_2(n) = 0'1 \cdot X_2(n-1) + u(n)$$

$$Y_2(n) = 0'1 \cdot Y_2(n-1) + v(n)$$

$$3. X_3(n) = 0'7 \cdot X_3(n-1) + u(n)$$

$$Y_3(n) = 0'4 \cdot Y_3(n-1) + v(n)$$

$$4. X_4(n) = X_4(n-1) + u(n)$$

$$Y_4(n) = Y_4(n-1) + v(n)$$

$$5. X_5(n) = 0'5 \cdot X_5(n-1) + 0'4 \cdot u(n-1) \cdot u(n)$$

$$Y_5(n) = 0'5 \cdot Y_5(n-1) + 0'4 \cdot v(n-1) \cdot v(n)$$

$$6. X_6(n) = u(n-1) \cdot u(n-1) + u(n)$$

$$Y_6(n) = v(n-1) \cdot v(n-1) + v(n)$$

Cada serie generada es de tamaño 100000 y se han descartado los 200 primeros datos con el fin de eliminar el efecto del valor inicial.



Las series $u(n)$ y $v(n)$ son ruidos blancos aleatorios, independientes entre sí. Como se puede apreciar, todos estos pares de series son independientes ya que ninguna serie $Y_i(n)$ depende de $X_i(n)$, ni viceversa. Se ha elegido mayoritariamente para $X(n)$ e $Y(n)$ el mismo esquema de reproducción, pero de forma independiente, para comprobar la eficacia del método propuesto.

Obsérvese que los cuatro primeros modelos generados (pares de series (1), (2), (3) y (4)) se corresponden con procesos autorregresivos de orden 1 con distintos coeficientes, AR(1), en lo que el valor de la serie en un instante depende del valor que tomó en el instante anterior. Además, el cuarto es el único no estacionario. El modelo (5) presenta una parte autorregresiva de orden 1 y una parte producto de ruidos. Por último, el modelo (6), cada serie depende del cuadrado del ruido blanco en el instante anterior más ese mismo ruido blanco en ese instante.

En segundo lugar, se proponen seis pares de modelos dependientes, los cuales son:

$$7. X_7(n) = 0'9 \cdot X_7(n-1) + u(n)$$

$$Y_7(n) = 0'5 \cdot Y_7(n-1) + 0'3 \cdot X_7(n-1) + v(n)$$

$$8. X_8(n) = 0'7 \cdot X_8(n-1) + u(n)$$

$$Y_8(n) = 0'3 \cdot Y_8(n-1) + 0'5 \cdot X_8(n-1) \cdot Y_8(n-2) + v(n)$$

$$9. X_9(n) = 0'7 \cdot X_9(n-1) + u(n)$$

$$Y_9(n) = 0'5 \cdot Y_9(n-1) + 0'3 \cdot X_9(n-1) \cdot Y_9(n-2) + v(n)$$

$$10. X_{10}(n) = u(n)$$

$$Y_{10}(n) = 0'9 \cdot X_{10}(n-1) \cdot X_{10}(n-1) + v(n)$$

$$11. X_{11}(n) = u(n)$$

$$Y_{11}(n) = \frac{4}{|X_{11}(n-1)| + 1} + v(n)$$

$$12. X_{12}(n) = 0'5 \cdot X_{12}(n-1) + u(n)$$

$$Y_{12}(n) = u(n-1) \cdot u(n-1) + v(n)$$

En este caso, todos los pares de series son dependientes. Se puede ver como las series $Y_i(n)$ siempre tienen relación, es decir dependencia, de $X_i(n)$. Se ha elegido distintos tipos de dependencia (lineal, no lineal, interacción débil, etc.) para comprobar la eficacia del método propuesto.

En el modelo simulado (7) existe una relación lineal entre las series. En (8) y (9) hay dependencia por interacción débil, siendo la segunda más débil que la primera. En el par (10) se ve como existe una dependencia, pero en este caso no es lineal, concretamente



es cuadrática. En (11) la dependencia es de cociente no lineal y en (12) la dependencia de $Y(n)$ no es directamente con $X(n)$, sino que lo es con el ruido blanco que ayuda a formar $X(n)$.

Para realizar el análisis entre los distintos pares de series hemos variado los parámetros w (tamaño de ventana) y m (dimensión del embedding). El parámetro m tomará los valores 3 y 4. No se usará $m=5$ porque se necesitaría que la longitud de la serie fuese más extensa. El parámetro w tomará valores de 1000, 3000 y 5000. También en este caso no toma valores mayores porque tendríamos demasiadas subventanas de tamaño m (recordar que el test Ji-cuadrado es sensible al tamaño muestral). De esta forma podremos ver el efecto de aumentar el tamaño de la ventana móvil y el efecto de aumentar el embedding.

Para decidir si dos series son dependientes o independientes los criterios de aceptación y rechazo de la independencia varían según se modifique el parámetro m . Si $m = 3$, el valor crítico de nuestro contraste al 95% de confianza viene dado por una chi-cuadrado con 25 grados de libertad $((3! - 1) \cdot (3! - 1) = 5 \cdot 5)$ y vale 37'65. A su vez, si $m = 4$ el valor crítico de nuestro contraste al 95% tiene un valor de 583'61. Por tanto, si la medida de asociación o estadístico de contraste de una ventana móvil es menor de 37'65 (para $m = 3$) o 583'61 (con $m = 4$), aceptaremos la independencia para esa ventana, y si es mayor, aceptaremos la dependencia. Además, cuanto mayor sea la medida de asociación más fuerte es la asociación entre las series.

Sin embargo, para la correlación lineal de Pearson y Spearman, los criterios de aceptación de independencia cambian. Para ello, hemos usado los p-valores respectivos, de forma que si su valor está por encima de 0'05, aceptaremos la independencia para esa ventana, y si es menor, aceptaremos la dependencia. Estos criterios de aceptación son iguales, sin importar el valor que pueda adoptar la dimensión del embedding.

Las siguientes tablas recogen el porcentaje de “fallos” de cada uno de los estadísticos para una dimensión de embedding y tamaño de ventana concreto. Cuando se dice “fallo” se hace referencia a que el estadístico afirma de forma errónea la dependencia o independencia del par de series. Recordar que los seis primeros pares son independientes y los seis últimos dependientes.



Caso 1: $m = 3$, $w = 1000$

MODELO	Chi	CR	G-test	Spearman	Pearson
S_1	11	12	15	45	50
S_2	5	5	5	5	7
S_3	5	5	5	12	15
S_4	6	5	7	90	92
S_5	9	9	11	20	18
S_6	6	6	7	5	2
S_7	0	0	0	0	0
S_8	57	56	55	70	43
S_9	80	80	78	68	57
S_{10}	1	1	1	94	96
S_{11}	2	2	2	94	93
S_{12}	3	3	3	97	93

Caso 2: $m = 3$, $w = 3000$

MODELO	Chi	CR	G-test	Spearman	Pearson
S_1	18,18	18,18	18,18	51,52	51,52
S_2	12,12	12,12	12,12	3,03	6,06
S_3	0,00	0,00	0,00	18,18	15,15
S_4	6,06	6,06	9,09	96,97	93,94
S_5	3,03	3,03	3,03	24,24	24,24
S_6	6,06	6,06	6,06	9,09	9,09
S_7	0,00	0,00	0,00	0,00	0,00
S_8	3,03	3,03	3,03	69,70	36,36
S_9	48,48	48,48	42,42	63,64	54,55
S_{10}	0,00	0,00	0,00	87,88	90,91
S_{11}	0,00	0,00	0,00	96,97	96,97
S_{12}	0,00	0,00	0,00	93,94	87,88

Caso 3: $m = 3$, $w = 5000$

MODELO	Chi	CR	G-test	Spearman	Pearson
S_1	10	10	10	30	40
S_2	0	0	0	10	5
S_3	0	0	0	10	10
S_4	5	5	5	95	100
S_5	10	10	10	15	15
S_6	5	5	5	5	5
S_7	0	0	0	0	0
S_8	0	0	0	75	40
S_9	15	15	15	80	60
S_{10}	0	0	0	85	95
S_{11}	0	0	0	95	95
S_{12}	0	0	0	95	95



Caso 4: $m = 4$, $w = 3000$

MODELO	Chi	CR	G-test	Spearman	Pearson
S_1	15,15	15,15	51,52	51,52	51,52
S_2	0,00	0,00	21,21	3,03	6,06
S_3	6,06	6,06	39,39	18,18	15,15
S_4	3,03	3,03	51,52	96,97	93,94
S_5	9,09	6,06	45,45	24,24	24,24
S_6	18,18	21,21	33,33	9,09	0,00
S_7	0,00	0,00	0,00	0,00	0,00
S_8	0,00	0,00	0,00	69,70	36,36
S_9	42,42	42,42	12,12	63,64	54,55
S_{10}	0,00	0,00	0,00	87,88	90,91
S_{11}	0,00	0,00	0,00	96,97	96,97
S_{12}	0,00	0,00	0,00	93,94	87,88

Caso 5: $m = 4$, $w = 5000$

MODELO	Chi	CR	G-test	Spearman	Pearson
S_1	5	5	20	30	40
S_2	0	0	5	10	5
S_3	10	10	20	10	10
S_4	20	20	20	95	100
S_5	15	5	20	15	15
S_6	5	5	15	5	5
S_7	0	0	0	0	0
S_8	0	0	0	75	40
S_9	5	5	5	80	60
S_{10}	0	0	0	85	95
S_{11}	0	0	0	95	95
S_{12}	0	0	0	95	95

Lo primero que hay que destacar es la gran efectividad que presentan los estadísticos Ji-cuadrado, G-test y Cressie-Read en todos los casos de independencia.

En las series que no presentan dependencia (modelos del (1) al (6)), se observa que en la correlación de Spearman y Pearson suele fallar en un porcentaje alto de ocasiones para los modelos (1), (4) y (5). Esto es debido a que las series dependen en gran medida de su valor anterior, y al repetirse el esquema en ambas series, se equivocan al afirmar la dependencia del par.

En modelos (2) y (3) todos los estadísticos tienen una gran eficacia. Las correlaciones lineales de Spearman y Pearson no yerran con tanta frecuencia ya que las series dependen en menor medida de su propio valor anterior.

También en el modelo (6), todas las medidas parecen adecuadas para poder concluir independencia. Sólo conviene destacar que para $m = 4$ y $w = 3000$ en el estadístico G-test, no se obtienen los resultados deseados. Koehler demuestra que para el G-test, la



aproximación a X^2 es pobre cuando n/N es menor de 5 (donde n es el tamaño muestral y N el número de celdas de la tabla de contingencia). En nuestro caso $n = w + m - 1$ y $N = m! \times m!$, así que para $m = 4$ y $w = 3000$:

$$\frac{n}{N} = \frac{3000 - 4 + 1}{24 * 24} = \frac{2997}{576} = 5'2 \sim 5$$

La situación mejora al aumentar $w = 5000$, pues quedaría:

$$\frac{n}{N} = \frac{5000 - 4 + 1}{24 * 24} > 5$$

En los modelos (8) y (9) hay una dependencia por interacción débil (siendo especialmente débil la interacción del (9)). Para una dimensión de embedding y tamaño de ventana bajos, el porcentaje de fallo es muy alto en todos los estadísticos. Si aumentamos el tamaño de ventana móvil, se subsana el fallo para los estadísticos G-test, chi-cuadrado y Cressie-Read en la serie (8), que es de las dos la que presenta una dependencia más fuerte, pero en (9) sigue habiendo un porcentaje alto de fallos. Si se aumenta aún más w los resultados son óptimos. Se observa en este caso que aumentar la dimensión del embedding apenas mejora los resultados, pero si incrementamos el tamaño de la ventana deslizante se obtendrá unos resultados idóneos. Las correlaciones de Spearman y Pearson solo son eficaces para el par de series (7), que es el único que presenta una dependencia lineal fuerte, así que solo será útil usar estos estadísticos cuando exista una dependencia de este tipo

Hasta ahora, se ha estudiado la eficacia del método propuesto para detectar dependencias entre pares de series que, o bien eran dependientes a lo largo de toda su historia, o bien eran independientes. Con el fin de estudiar el comportamiento de las medidas propuestas en series que poseen tramos dependientes y tramos independientes, se considerará el siguiente par de series simuladas.

Se generaran series de tamaño 5000 para cada modelo del apartado anterior y las pegaremos alternando INDEP-DEP-INDEP-DEP ... Se tomarán, por ejemplo, los 5000 primeros datos de cada serie simulada en el apartado anterior.

$$X_{conc}(n) = X_1(n) \cup X_7(n) \cup X_2(n) \cup X_8(n) \cup X_3(n) \cup X_9(n) \cup X_4(n) \cup X_{10}(n) \\ \cup X_5(n) \cup X_{11}(n) \cup X_6(n) \cup X_{12}(n)$$

$$Y_{conc}(n) = Y_1(n) \cup Y_7(n) \cup Y_2(n) \cup Y_8(n) \cup Y_3(n) \cup Y_9(n) \cup Y_4(n) \cup Y_{10}(n) \cup Y_5(n) \\ \cup Y_{11}(n) \cup Y_6(n) \cup Y_{12}(n)$$

Se han generado en total 60000 datos para cada serie. Recordar que los seis primeros modelos son independientes y los seis últimos dependientes.

Para realizar el análisis hemos variado nuevamente los parámetros w (tamaño de ventana) y m (dimensión del embedding). El parámetro m tomará los valores 3 y 4 y el parámetro w tomará valores de 1000, 3000 y 5000. De esta forma podremos ver el efecto de aumentar el tamaño de la ventana móvil y el efecto de aumentar el embedding.

Igual que en el caso anterior, usaremos que para $m = 3$, la región de rechazo viene determinada por el valor 37'65 y para $m = 4$ por el valor 583'61.

Las siguientes gráficas recogen los resultados de cada uno de los estadísticos para una dimensión de embedding y tamaño de ventana concreto. Se representarán en el primer gráfico los estadísticos chi-cuadrado, G-test y Cressie-Read, y en el segundo los p-valores las correlaciones de Pearson y Spearman. Para distinguir los estadísticos de cada gráfica, se ha representado a cada uno de ellos de un color distinto: en las primeras gráficas Chi-cuadrado de negro, CR será azul y G-test rojo; en las segundas gráficas Spearman será la línea negra y Pearson la azul. La línea roja discontinua es la que nos limitará las regiones de aceptación y rechazo de la independencia al 95% de confianza. Como se ha dicho anteriormente, si en las primeras gráficas el valor del estadístico se sitúa por debajo de la línea roja discontinua, se aceptará la independencia, mientras que si se sitúa por debajo en las segundas gráficas (p-valores) se aceptará la dependencia.

Los resultados se mostrarán en las siguientes gráficas:

Caso 1: $m = 3$, $w = 1000$

Los 5 primeros valores de la gráfica son los resultados de los estadísticos para las cinco primeras ventanas móviles que sabemos que presentan independencia (son los 5000 primeros datos del modelo (1)), los 5 siguientes presentan dependencia (son los valores de las series del modelo (7)), y así sucesivamente.

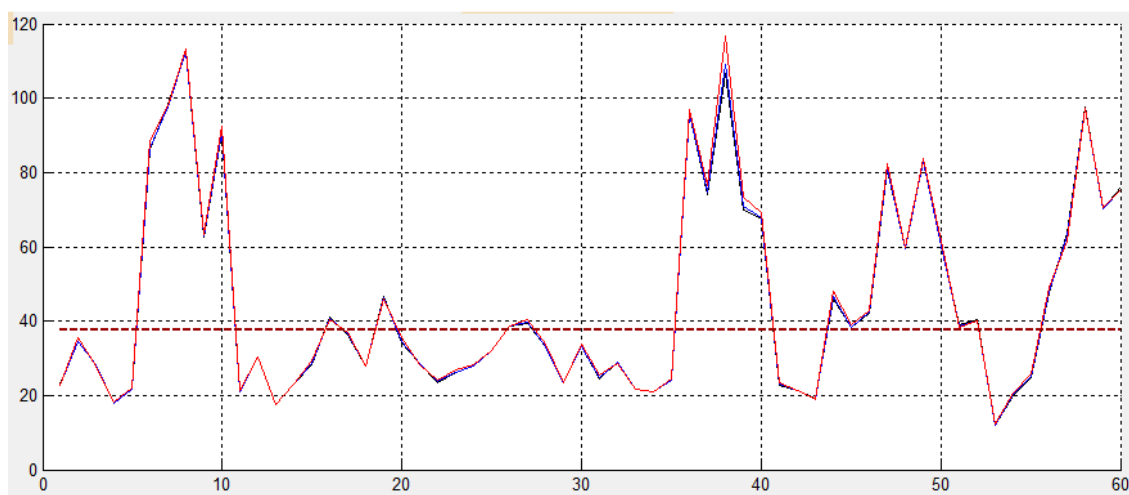


Figura 4.1: Medidas de asociación para serie concatenada, caso 1

La detección de la dependencia es bastante buena, menos en las ventanas móviles del 16-20 y del 26-30. En esos tramos las ventanas móviles contenían los datos de los

modelos (8) y (9), que como se ha visto con anterioridad, para esta dimensión del embedding y tamaño de ventana no detecta la dependencia con mucha eficacia.

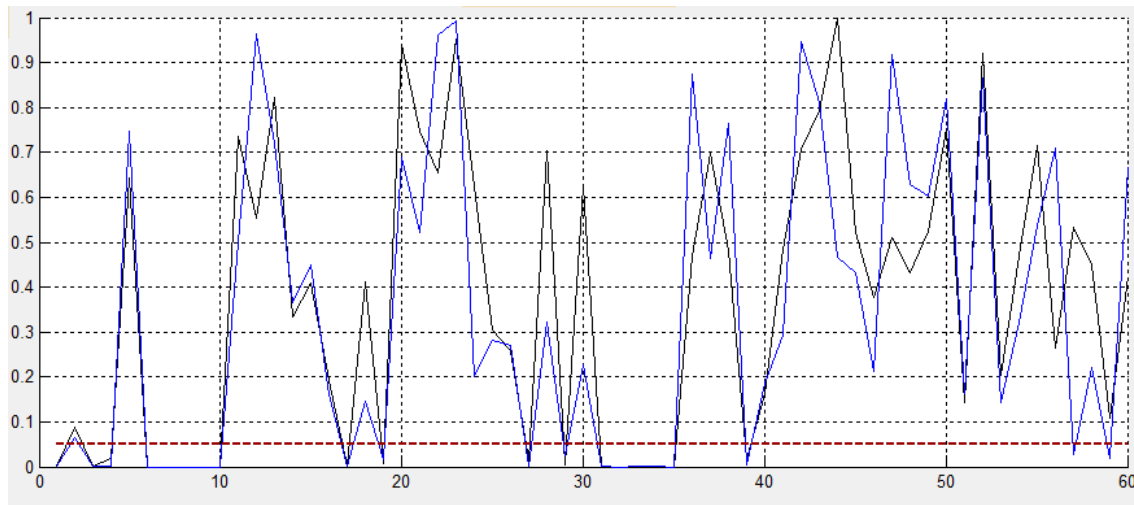


Figura 4.2: P-valores de Pearson y Spearman para serie concatenada, caso 1

Se aprecia que para las correlaciones de Pearson o Spearman solo se detecta con eficacia los tramos 6-10, 11-15, 21-25, 41-45 y 51-55 cuyos datos se corresponden con los pares de series (7), (2), (3), (5) y (6) respectivamente. En el apartado anterior se observa que son exactamente esas series donde éstos estadísticos presentan una mayor fiabilidad para la detección de dependencia.

Caso 2: $m = 3$, $w = 3000$

Este caso es especialmente importante, ya que se podrá analizar qué ocurre en las transiciones. Como se cogían los 5000 primeros datos de cada serie, al tener un tamaño de ventana de 3000, la segunda ventana móvil contendrá 2000 datos de una serie independiente y 1000 datos de una serie dependiente, y así sucesivamente.

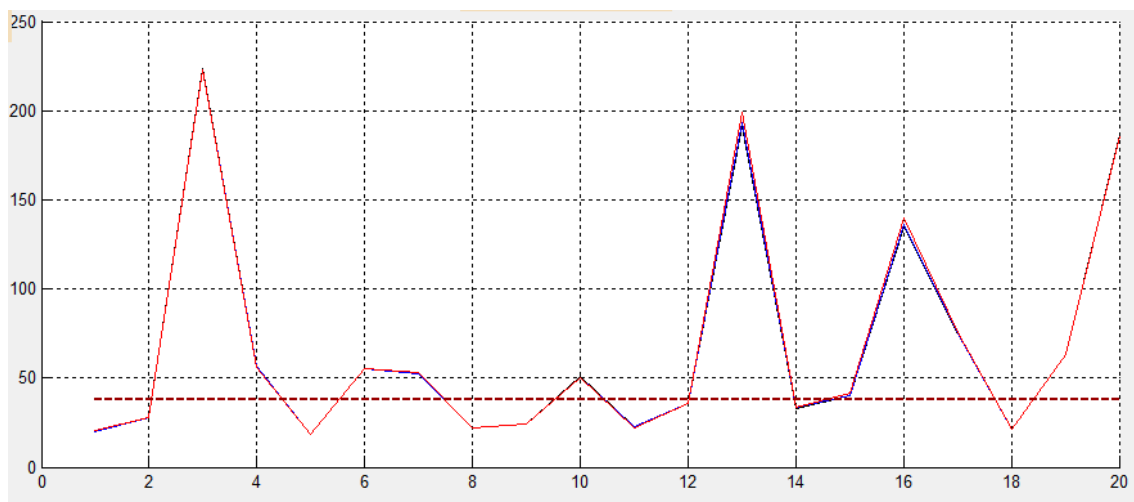


Figura 4.3: Medidas de asociación para serie concatenada, caso 2

La gráfica muestra que el segundo dato, las series presentan independencia. Esto es normal ya que la mayoría de datos analizados son independientes. De los 20 valores (ventanas móviles), tan solo tres muestran de forma “errónea” la existencia o no de dependenci: las ventanas 4, 9 y 15. En el dato 4 se compararon 1000 datos del par (7) y 2000 del (2). Predominan los datos independientes, de forma que debería salir independencia, mientras que los estadísticos nos informan que existe una dependencia, eso sí, muy débil, que es debido a la gran dependencia que presentan los datos del par de series (7). La ventana 9 contiene 1000 valores del modelo (3) y 2000 del (9), por lo que debería predominar la dependencia, pero las series (9) necesitaban un tamaño de ventana mayor para detectar esa dependencia. La última ventana anómala es la 15, que solo tiene valores del par de series (5), que es independiente, aunque la dependencia detectada es muy débil.

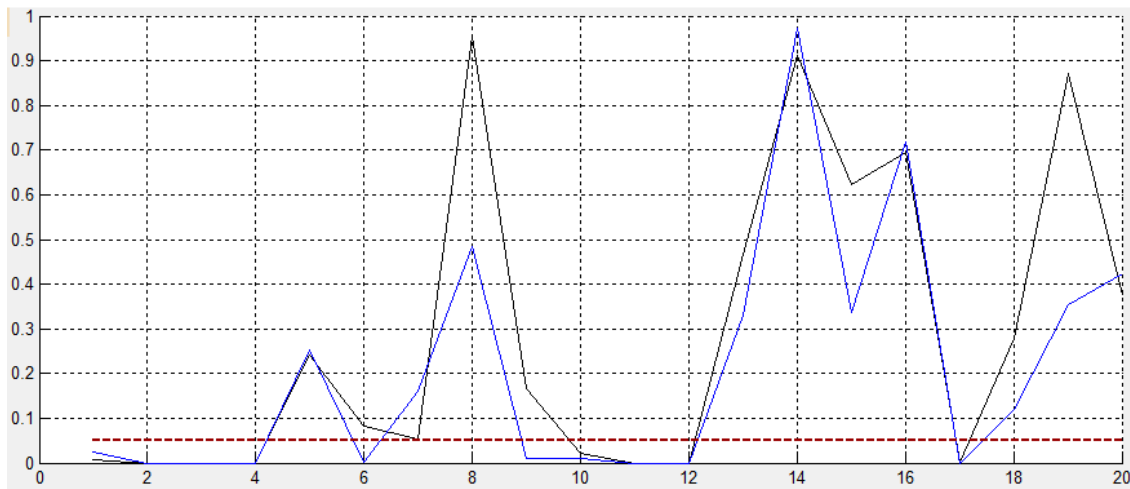


Figura 4.4: P-valores de Pearson y Spearman para serie concatenada, caso 2

Para los estadísticos de Pearson y Spearman aproximadamente la mitad de las conclusiones de dependencia son erróneas.

Caso 3: $m = 3$, $w = 5000$

Para estos valores de m y w se da el caso que cada ventana deslizante coincide con el un tramo de cada modelo simulado. Es decir, la gráfica debe mostrar independencia-dependencia de forma alternada.

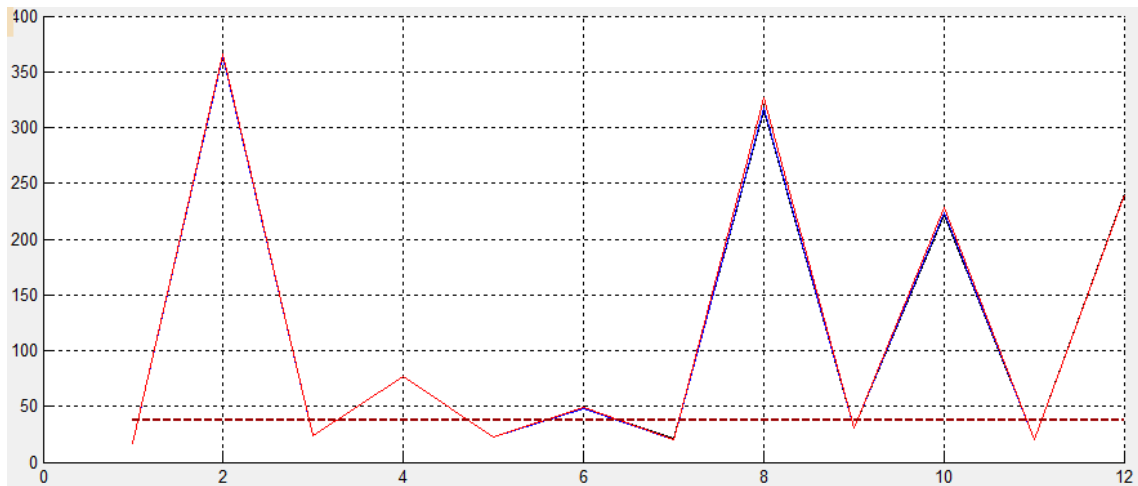


Figura 4.5: Medidas de asociación para serie concatenada, caso 3

Se observa en el gráfico que efectivamente los estadísticos predicen sin ningún tipo de fallo la independencia-dependencia.

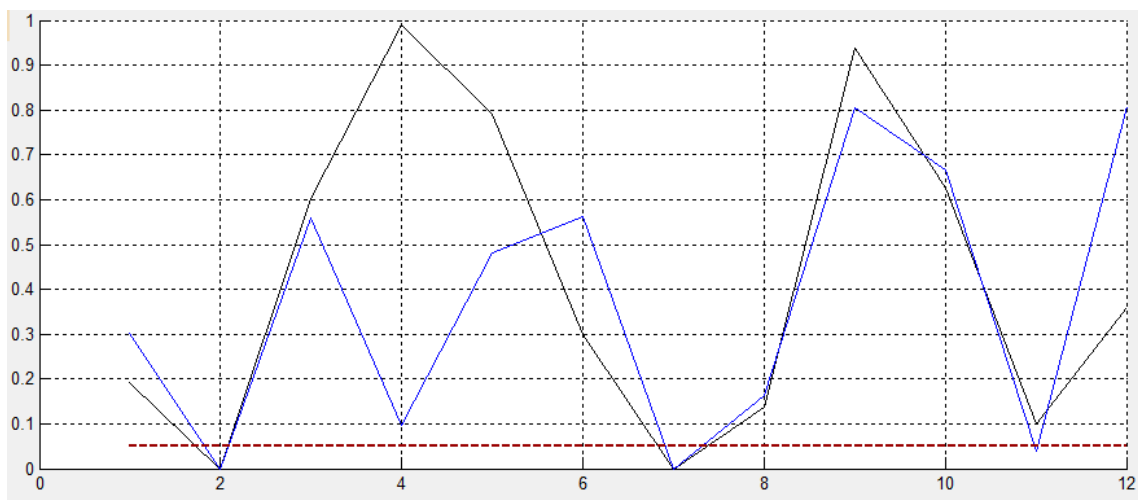


Figura 4.6: P-valores de Pearson y Spearman para serie concatenada, caso 3

Una vez más, se aprecia que las correlaciones de Spearman y Pearson solo detectan la independencia o dependencia con acierto en la mitad de los casos.

Caso 4: $m = 4$, $w = 3000$

En este caso, vamos a estudiar de nuevo qué sucede en las transiciones pero se ha aumentado la dimensión del embedding.

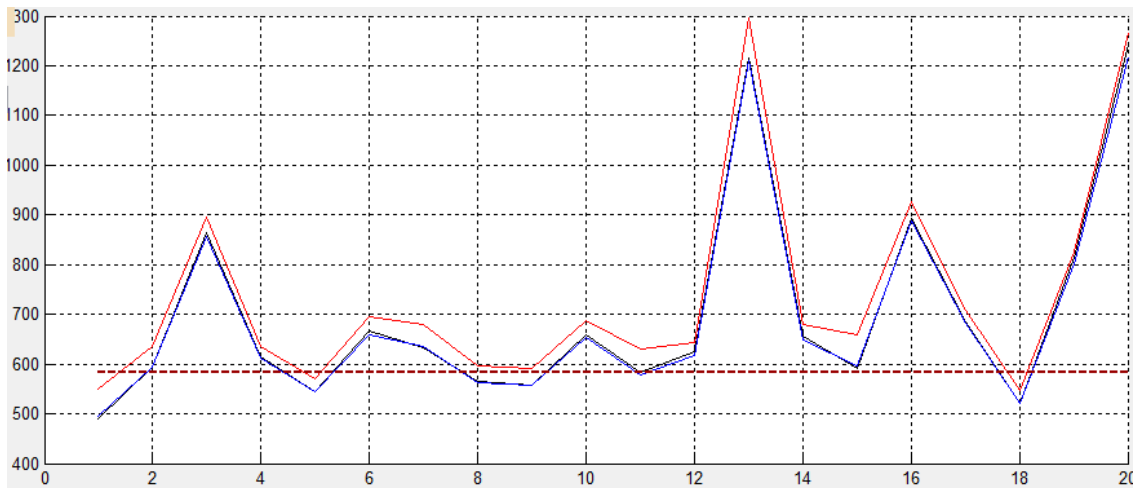


Figura 4.7: Medidas de asociación para serie concatenada, caso 4

Lo primero que se observa es que los resultados para el G-test son siempre un poco superiores a los de chi-cuadrado y Cressie-Read. Esto se debe a los valores que toman m y w que, según Koehler, hacen que este estadístico sea poco fiable con estos parámetros. Para los otros estadísticos las únicas ventanas que varían son 2, 12 y 15 con respecto al caso $m = 3$. La ventana 2 nos dice que las series presentan una dependencia muy débil, y aunque hay más tramo de datos que presentan independencia, los 1000 datos del par (7) tienen una gran dependencia. Lo mismo para la ventana 12. En la ventana 15 se corrige el fallo que había con $m = 3$ y nos indica que hay dependencia.

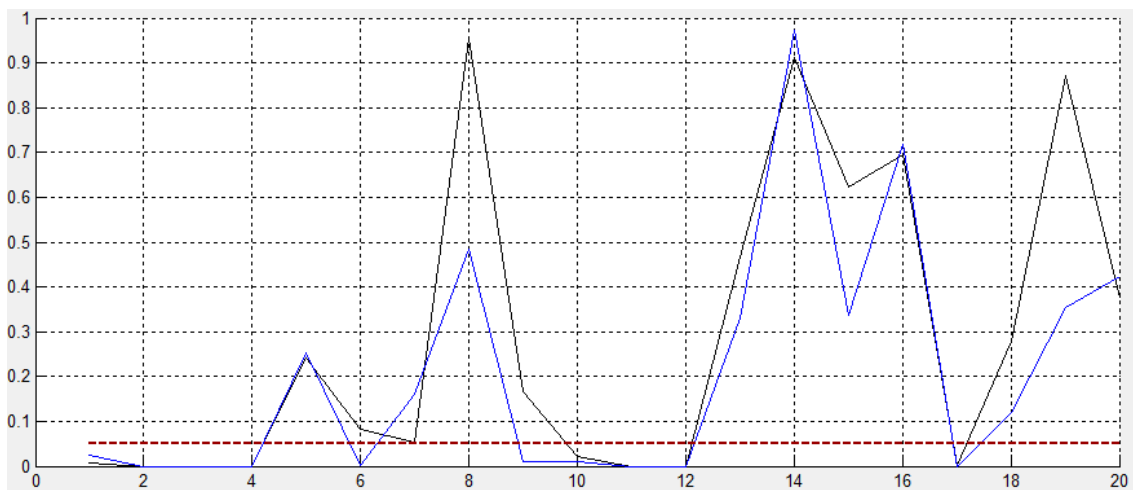


Figura 4.8: P-valores de Pearson y Spearman para serie concatenada, caso 3

Caso 5: $m = 4$, $w = 5000$

Para este valor de w se da el caso que cada ventana deslizante coincide con el tramo de par de series unido. Es decir, la gráfica debe mostrar independencia-dependencia de forma alternada. Es la misma situación que el caso 3 pero aumenta la dimensión del embedding.



Figura 4.9: Medidas de asociación para serie concatenada, caso 5

Al igual que ocurría con $m = 3$, los estadísticos predicen sin ningún tipo de fallo la independencia-dependencia, aunque en este caso el valor de G-test es un poco más elevado que el de los demás.

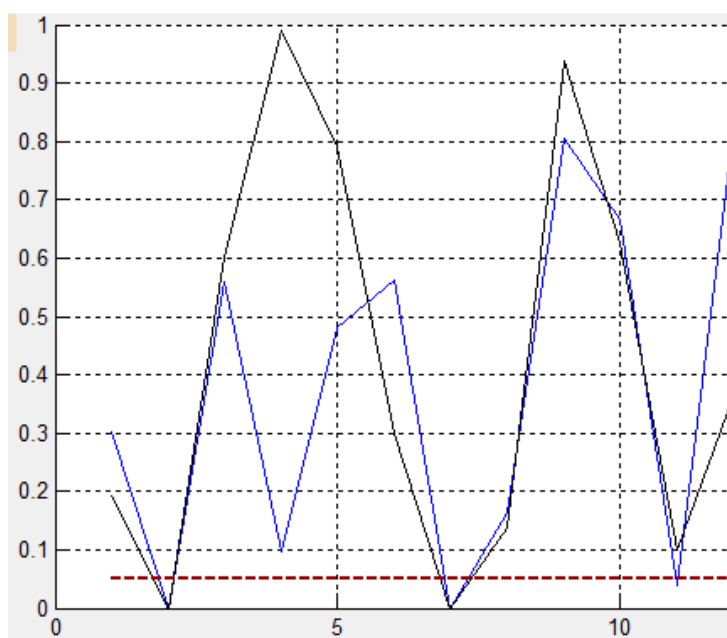


Figura 4.10: P-valores de Pearson y Spearman para serie concatenada, caso 5



4.2. ESTUDIO DE LA DEPENDENCIA ANTE LA PRESENCIA DE ESTACIONALIDAD

En esta parte se analizará cómo afecta la presencia de estacionalidad en el estudio de la dependencia de series. Consideraremos 3 modelos (pares de series) con componente estacional de periodo 24 de la siguiente forma:

$$1. \quad X(n) = X(n - 24) + u(n)$$

$$Y(n) = Y(n - 24) + v(n)$$

$$2. \quad X(n) = 0'9 \times X(n - 24) + u(n)$$

$$Y(n) = 0'9 \times Y(n - 24) + v(n)$$

$$3. \quad X(n) = 0'5 \times X(n - 24) + u(n)$$

$$Y(n) = 0'5 \times Y(n - 24) + v(n)$$

En los tres casos, las series $X(n)$ e $Y(n)$ son independientes, aunque ambas se han generado con el mismo esquema. La diferencia está en que el grado de determinismo de los modelos 1 y 2 es significativamente mayor que en el modelo 3. Es decir, en los modelos 1 y 2, la estacionalidad es la componente dominante mientras que en el modelo 3, el ruido pasa a ser la componente dominante.

Se han generado 100000 datos para cada serie y se analizan dichas series con los parámetros que se han utilizado en los anteriores apartados. Las siguientes tablas recogen el porcentaje de “fallos” de cada uno de los estadísticos para una dimensión de embedding y tamaño de ventana concreto. Cuando se dice “fallo” se hace referencia a que el estadístico afirma de forma errónea la dependencia o independencia del par de series.

Caso 1: $m = 3$, $w = 1000$

MODELO	Chi	CR	G-test	Spearman	Pearson
S_1	100	100	100	82	70
S_2	93	93	94	46	50
S_3	15	14	15	16	15

Obsérvese que los estadísticos indican que los modelos 1 y 2 presentan dependencia, cuando no es cierto. Se debe a la presencia de componente estacional, siendo además la componente dominante. Sin embargo, el modelo 3 sale independiente, lo que se corresponde con la realidad, porque en este caso la componente aleatoria (ruido) es la dominante.



Caso 2: $m = 3$, $w = 3000$

MODELO	Chi	CR	G-test	Spearman	Pearson
S_1	100,00	100,00	100,00	87,88	78,79
S_2	87,88	87,88	90,91	24,24	27,27
S_3	12,12	12,12	12,12	12,12	24,24

Caso 3: $m = 3$, $w = 5000$

MODELO	Chi	CR	G-test	Spearman	Pearson
S_1	100	100	100	90	85
S_2	90	90	90	35	35
S_3	25	20	20	25	15

Caso 4: $m = 4$, $w = 3000$

MODELO	Chi	CR	G-test	Spearman	Pearson
S_1	100,00	100,00	100,00	87,88	78,79
S_2	100,00	100,00	100,00	24,24	27,27
S_3	3,03	3,03	18,18	12,12	24,24

Caso 5: $m = 4$, $w = 5000$

MODELO	Chi	CR	G-test	Spearman	Pearson
S_1	100	100	100	90	85
S_2	100	100	100	35	35
S_3	20	20	25	25	15

Como se puede observar, los resultados apenas varían al cambiar la dimensión del embedding o el tamaño de la ventana.

Para evitar las “falsas” dependencias debidas a la presencia de componente estacional, se requiere eliminar dicha componente de las series en estudio antes de estudiar su asociación. A modo de ejemplo, consideremos los datos del modelo 1 y le extraemos su componente estocástica (eliminamos la estacionalidad). Para ello, basta con considerar las series

$$Z_x = X(n) - X(n - 24)$$

$$Z_y = Y(n) - Y(n - 24)$$

Si calculamos los estadísticos usuales para las nuevas series sin componente estacional, el porcentaje de fallos se ha reducido considerablemente, como se observa en la siguiente tabla.



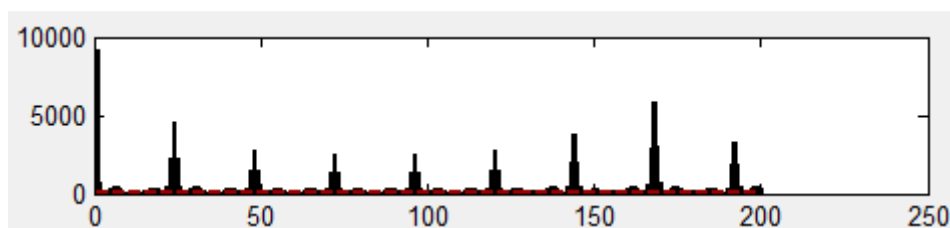
MODELO	Chi	CR	G-test	Spearman	Pearson
$m = 3; w = 1000$	13	13	14	10	8
$m = 3; w = 3000$	15,15	15,15	15,15	9,09	6,06
$m = 3; w = 5000$	10	10	10	10	15
$m = 4; w = 3000$	6,06	6,06	24,24	9,09	6,06
$m = 4; w = 5000$	10	10	10	10	15

Por tanto, al desestacionalizar el modelo 1, se observa los estadísticos empleados indican independencia entre las series Z_x y Z_y , que es lo correcto. El cambio de la dimensión del embedding y el del tamaño de ventana apenas afecta.

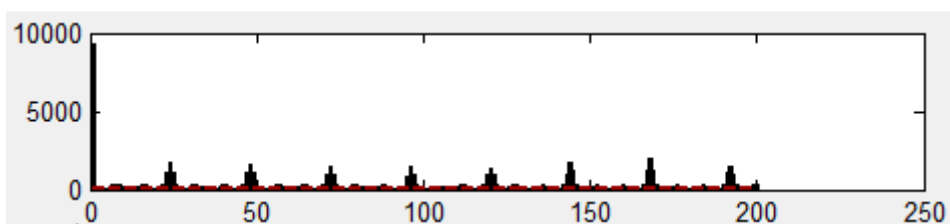
El interés de estudiar qué sucede ante la presencia de estacionalidad se debe a que las series de precios de electricidad presentan componente estacional de periodo semanal (168 horas) y diario (24 horas), siendo además la componente estacional predominante frente al ruido, así que estas series de precios deberían tratarse de la forma adecuada.

Para ilustrar el efecto de la componente estacional en el estudio de la dependencia, usaremos los datos de 2 mercados muy distintos en cuanto a funcionamiento y alejados geográficamente, OMEL y Ontario. Primero haremos un autocorrelograma de cada uno de los mercados:

OMEL



Ontario



Se observa que existe una clara componente estacional, tanto diaria como semanal.

En segundo lugar se va a calcular las medidas de asociación Ji-cuadrado, G-test y CR, p-valores de Pearson y Spearman, para los datos originales, es decir, CON componente estacional.

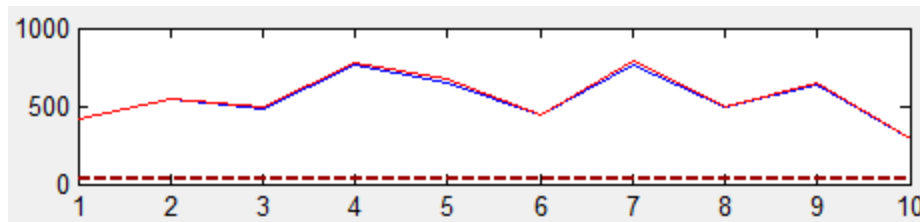


Figura 4.11: Medidas de asociación entre OMEL y Nordpool originales

Se observa que existe una clara dependencia entre ambos mercados.

Finalmente, eliminaremos la componente estacional de cada serie (con diferenciación estacional semanal) y volveremos a calcular las medidas de asociación:

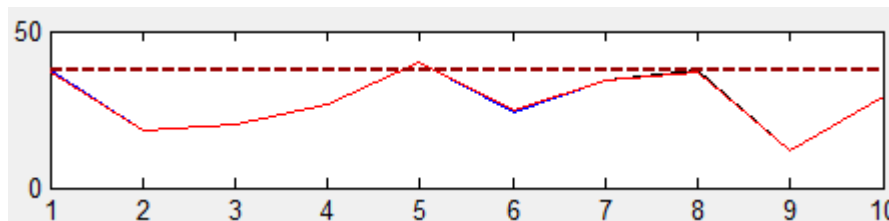


Figura 4.12: Medidas de asociación entre OMEL y Nordpool desestacionalizados

Se observa que ya en este caso las medidas de asociación de Ji-cuadrado, G-test y Cressie-Read muestran independencia.

4.3. IDENTIFICACIÓN DEL RETARDO QUE PROPORCIONA MÁXIMA DEPENDENCIA

En esta sección se pretende analizar cómo identificar el retardo en el tiempo que proporciona mayor dependencia funcional entre las series. Para ello se definirá previamente qué es un correlograma cruzado.

Consideremos dos procesos estacionarios x_t e y_t , para $t = 0, \pm 1, \pm 2, \dots$. La covarianza cruzada de orden k entre x_t e y_t está definida como (por ejemplo, Box & Jenkins 1976, Wei 1990):

$$\gamma_{xy}(k) = E[(X_t - \mu_x)(y_{t+k} - \mu_y)]$$

para $k = 0, \pm 1, \pm 2, \dots$



$\gamma_{xy}(k)$ es conocida como la función de covarianza cruzada entre x_t e y_t , que depende del retardo k

La estandarización de $\gamma_{xy}(k)$ produce la función de correlación cruzada (FCC)

$$\rho_{xy}(k) = \gamma_{xy}(k) / (\sigma_x \sigma_y)$$

para $k = 0, \pm 1, \pm 2, \dots$, donde σ_x y σ_y son las desviaciones estándar de los procesos x_t e y_t .

La FCC mide no solamente la fortaleza de la relación, sino también su dirección. Esta última propiedad es útil para identificar variables causales. Por esta razón, es importante examinar la FCC tanto para los valores positivos de k como para los negativos. Para valores negativos de k , la FCC describe la influencia lineal de los valores pasados de y_t sobre x_t . Para valores positivos de k , la FCC indica la influencia lineal de los valores pasados de x_t sobre y_t . El gráfico de la FCC contra k , llamado correlograma cruzado, es útil para visualizar estas relaciones.

Dada una realización de n periodos del proceso estacionario bivalente x_t, y_t , la FCC se estima de la siguiente manera:

$$\hat{\rho}_{xy}(k) = \frac{\hat{\gamma}_{xy}(k)}{\hat{\sigma}_x \hat{\sigma}_y}$$

para $k = 0, \pm 1, \pm 2, \dots$, donde

$$\begin{aligned} \hat{\gamma}_{xy}(k) &= \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(y_{t+k} - \bar{y}), \quad \text{si } k \geq 0 \\ &= \frac{1}{n} \sum_{t=1-k}^n (x_t - \bar{x})(y_{t+k} - \bar{y}), \quad \text{si } k < 0 \end{aligned}$$

Y donde $\hat{\sigma}_x = [\gamma_{xx}(0)]^{1/2}$, $\hat{\sigma}_y = [\gamma_{yy}(0)]^{1/2}$, \bar{x} e \bar{y} son las desviaciones estándar y las medias muestrales de las series x_t e y_t , respectivamente.

A continuación se expondrá un ejemplo de la aplicación de los correlogramas cruzados. Se va a analizar como va a ser el correlograma cruzado de precio de la electricidad con el consumo de ésta del mercado eléctrico de Canadá (Ontario).

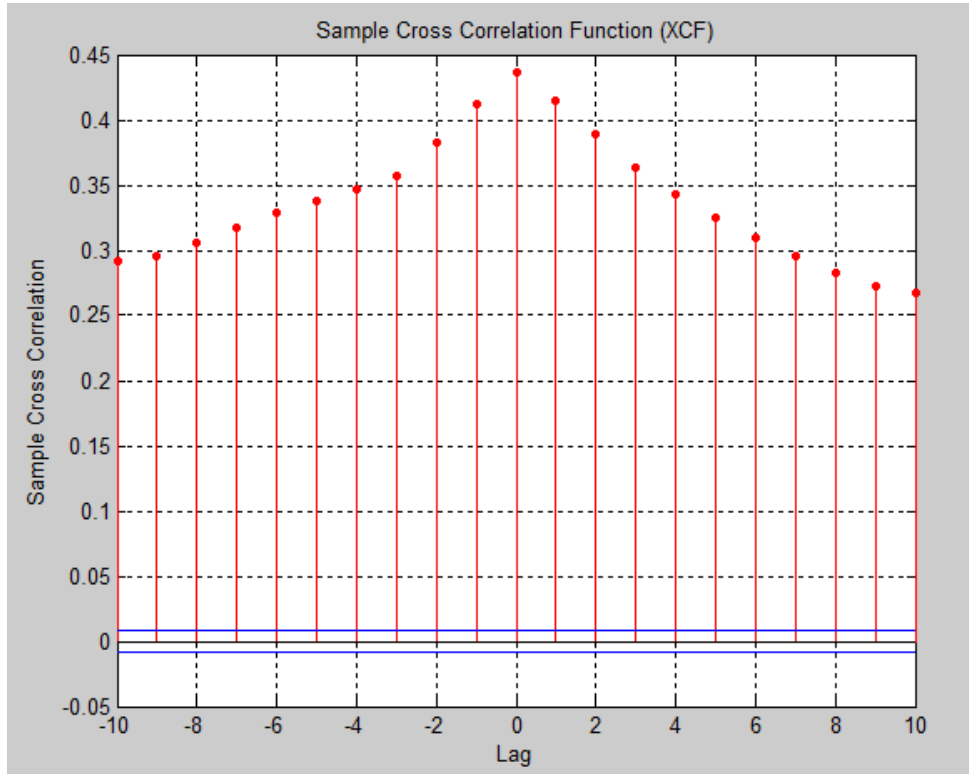


Figura 4.13: Autocorrelograma de precio y consumo eléctrico en Ontario

Se observa que la mayor relación de dependencia lineal existente entre el precio de la electricidad y el consumo es en el instante 0, es decir, el precio de la electricidad en un instante depende en mayor medida del consumo que se está haciendo en ese mismo instante.

Destacar que el programa que se ha diseñado para este punto proporciona el gráfico equivalente a un correlograma cruzado (que podríamos llamar dependiograma cruzado) pero referido a las medidas de asociación que se han propuesto en lugar de a la correlación de Pearson. Como se hizo anteriormente, la eficacia del método ha sido contrastada mediante algunos modelos simulados que describimos a continuación.

En la primera parte de este apartado, la serie $X(n)$ es un ruido gaussiano, mientras que la serie $Y(n)$ depende de la $X(n)$ de distintas formas (modelos $Y_1(n)$ hasta $Y_7(n)$):

1. $Y_1(n) = 0'8 \cdot X(n) + v(n)$
2. $Y_2(n) = 0'8 \cdot X(n - 1) + v(n)$
3. $Y_3(n) = 0'8 \cdot X(n - 10) + v(n)$
4. $Y_4(n) = 0'8 \cdot X(n - 10) \cdot X(n - 10) + v(n)$
5. $Y_5(n) = 0'3 \cdot X(n - 1) + 0'5 \cdot X(n - 10) + v(n)$
6. $Y_6(n) = 0'5 \cdot X(n - 1)^2 + 0'3 \cdot X(n - 10)^2 + v(n)$
7. $Y_7(n) = 0'5 \cdot X(n - 1)^3 + 0'3 \cdot X(n - 10)^3 + v(n)$

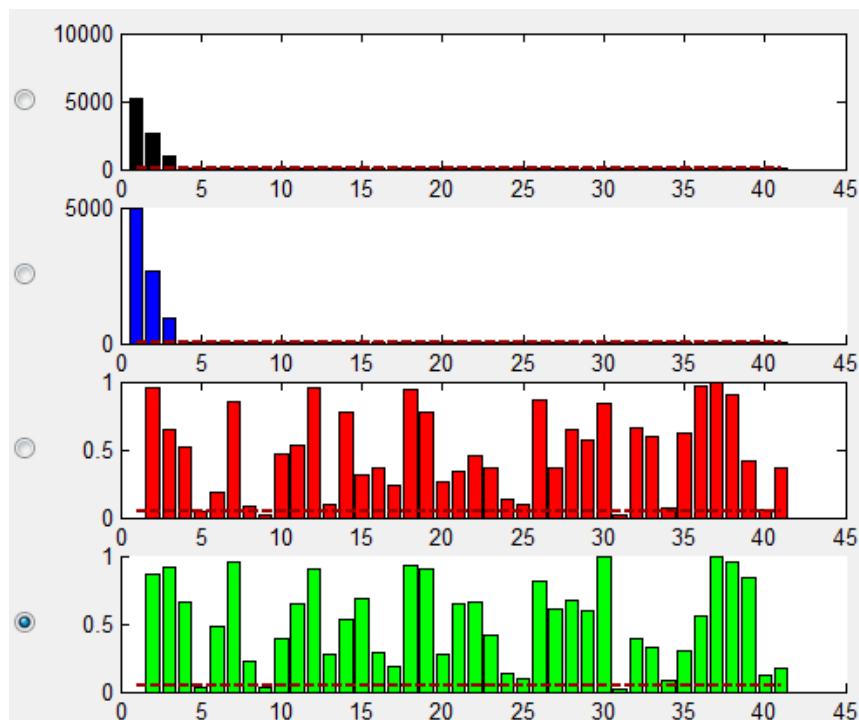


La longitud de cada una de las series es de 10000 datos. En este caso el tamaño de ventana será siempre el mismo y tendrá el mismo valor que la longitud de los datos. Lo que sí se modificará será la dimensión del embedding. Se realizarán las comparaciones para $m = 3$ y $m = 4$, de esta forma podremos ver el efecto de aumentar el embedding. Se mostrará un rango de retardos desde 0 hasta 40.

Los criterios de aceptación y rechazo de la independencia varían según se modifique el parámetro m , de la misma forma que en los apartados anteriores de este proyecto, de forma que para $m = 3$, el valor crítico de nuestro contraste vale 37'65 y para $m = 4$ el valor es de 583'61. Para la correlación lineal de Pearson y Spearman, los criterios de aceptación de independencia cambian, de forma que si su valor está por encima de 0'05, aceptaremos la independencia para esa ventana, y si es menor, aceptaremos la dependencia.

Se representarán los retardos para cuatro estadísticos. Para facilitar su visión, cada uno de ellos tendrá un color distinto. La grafica superior, con color negro, se corresponde con el chi-cuadrado. La azul será el G-test. No se mostrará el Cressie-Read, porque, como se ha visto, sus resultados son casi siempre similares al chi-cuadrado. Los p-valor Spearman será el correlograma de color rojo y el último, el verde, será el p-valor de Pearson.

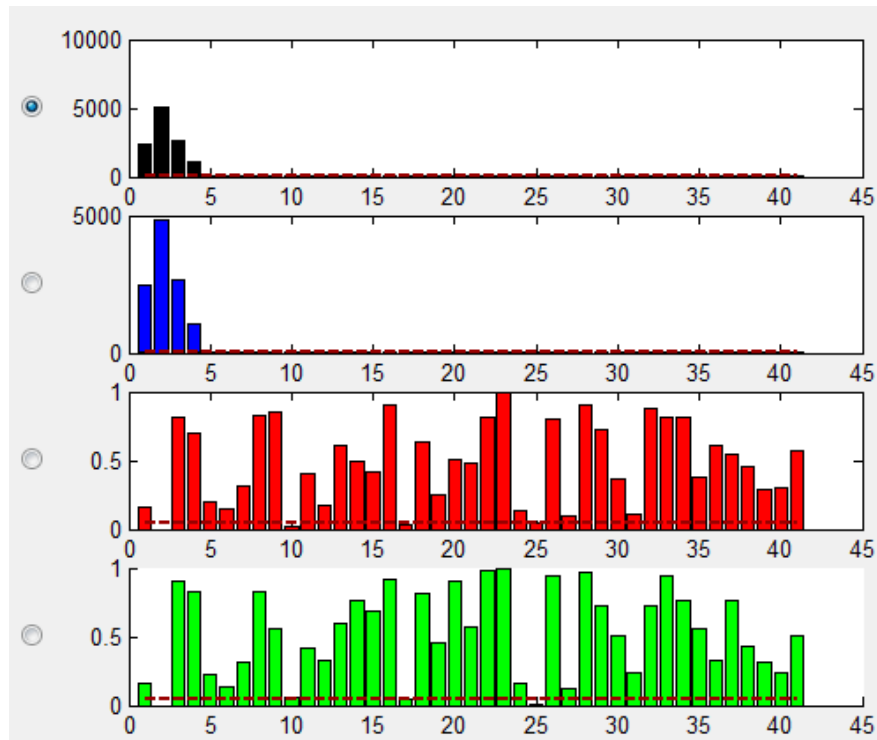
Caso 1: $m = 3$, $X(n) - Y_1(n)$, donde $Y_1(n) = 0'8 \cdot X(n) + v(n)$



Todos los estadísticos muestran la máxima dependencia con retardo cero, de forma que los resultados son los deseados. Los estadísticos de correlación de Pearson y Spearman también la detectan ya que la dependencia del modelo es lineal.

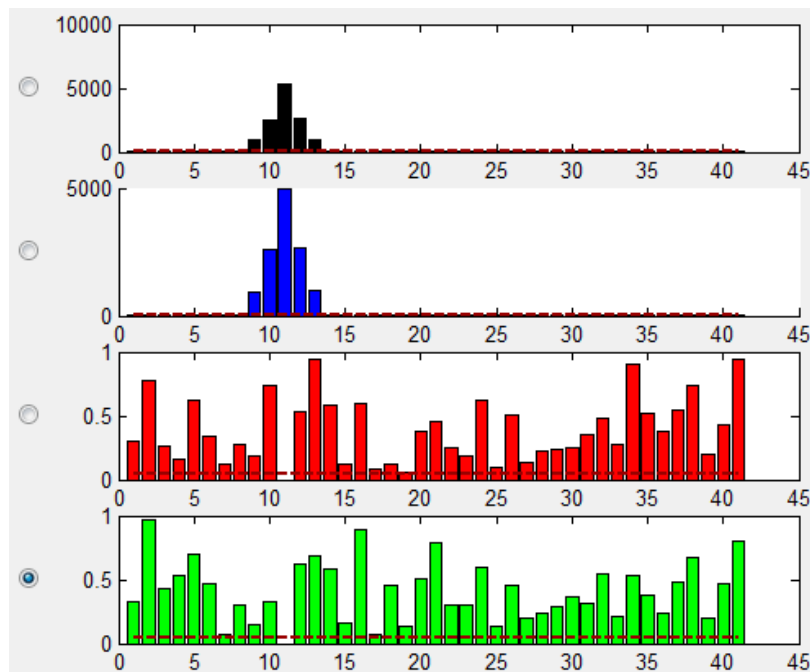


Caso 2: $m = 3$, $X(n) - Y_2(n)$, donde $Y_2(n) = 0'8 \cdot X(n - 1) + v(n)$

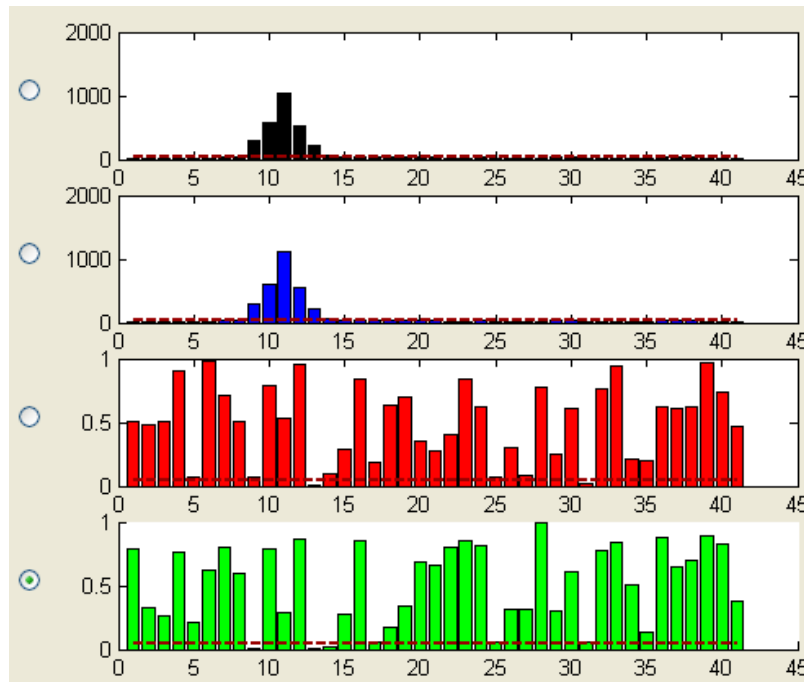


Este caso es idéntico al Caso 1, pero en este caso la máxima dependencia se encuentra en el retardo 1.

Caso 3: $m = 3$, $X(n) - Y_3(n)$, donde $Y_3(n) = 0'8 \cdot X(n - 10) + v(n)$

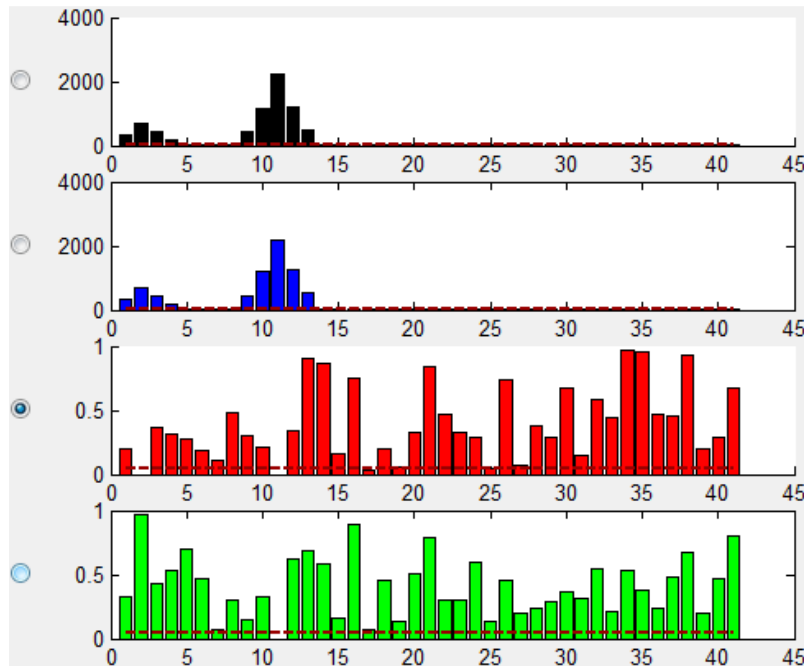


Caso 4: $m = 3$, $X(n) - Y_4(n)$, donde $Y_4(n) = 0'8 \cdot X(n - 10) \cdot X(n - 10) + v(n)$



Igual caso que los anteriores, pero con dependencia cuadrática. Destacar que en este caso las correlaciones de Spearman y Pearson fallan en la detección de la dependencia.

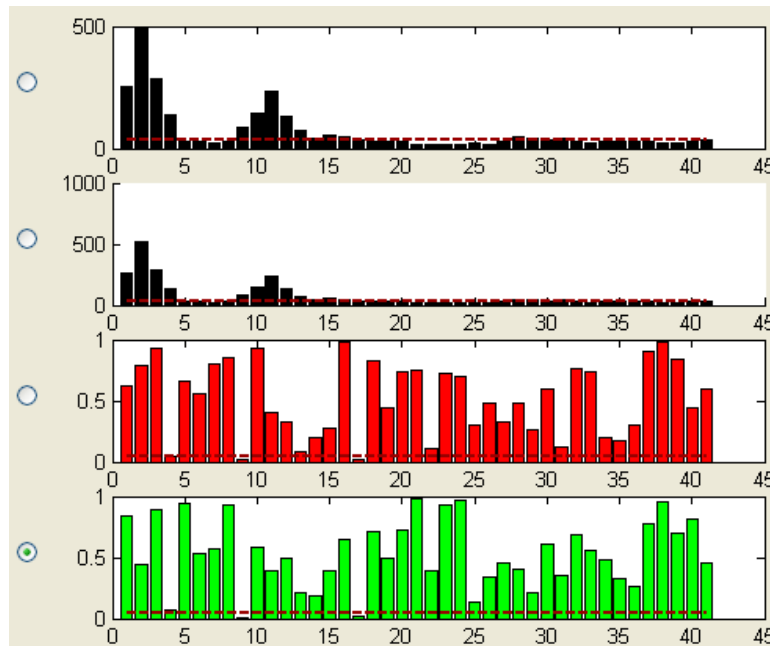
Caso 5: $m=3$, $X(n) - Y_5(n)$, donde $Y_5(n) = 0'3 \cdot X(n - 1) + 0'5 \cdot X(n - 10) + v(n)$



Ahora las máximas dependencias se encuentran en los retardos 1 y 10, aunque es mayor en 10. En este caso la dependencia en el retardo en la correlación de Pearson no es detectada.

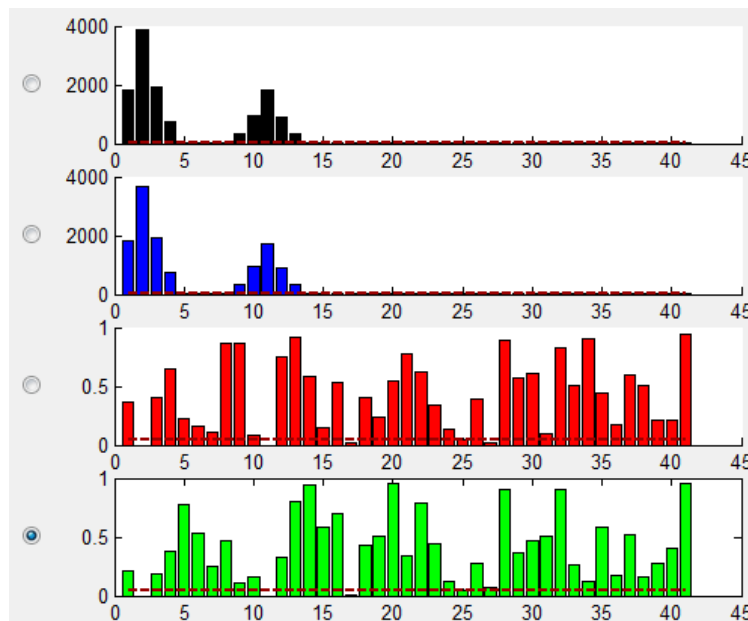


Caso 6: $m=3$, $X(n) - Y_6(n)$, donde $Y_6(n) = 0'5 \cdot X(n-1)^2 + 0'3 \cdot X(n-10)^2 + v(n)$



Igual caso que el anterior, pero con dependencias cuadráticas. Solo son detectadas con éxito en los estadísticos de Ji-cuadrado y G-test, mientras que en las correlaciones de Pearson y Spearman los resultados son erróneos.

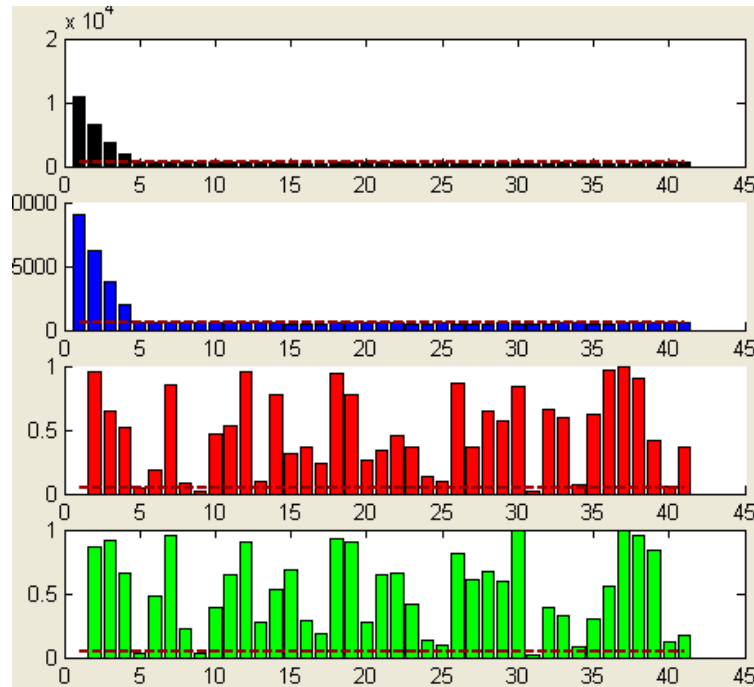
Caso 7: $m = 3$, $X(n) - Y_7(n)$, donde $Y_7(n) = 0'5 \cdot X(n-1)^3 + 0'3 \cdot X(n-10)^3 + v(n)$



Es un caso igual al 5 pero con dependencia cúbica. Todos los estadísticos muestran la dependencias con éxito.

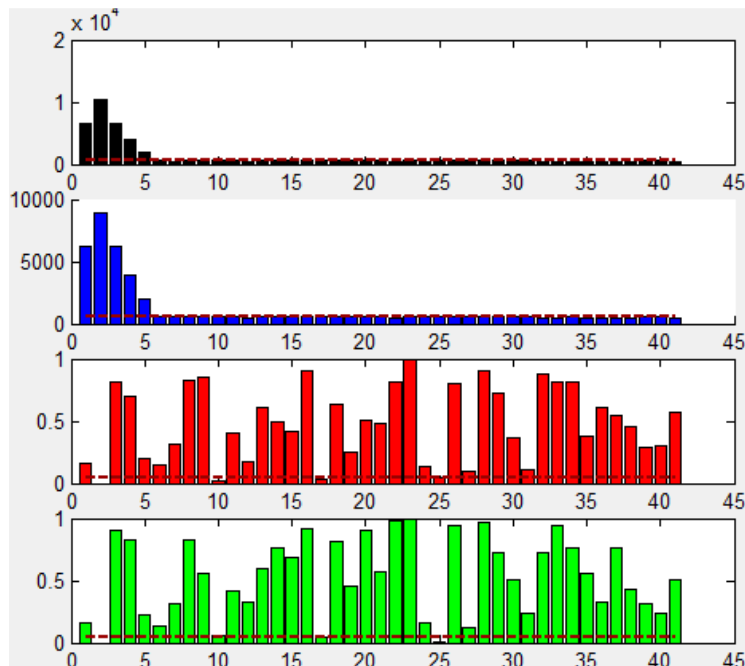


Caso 8: $m = 4$, $X(n) - Y_1(n)$, donde $Y_1(n) = 0'8 \cdot X(n) + v(n)$

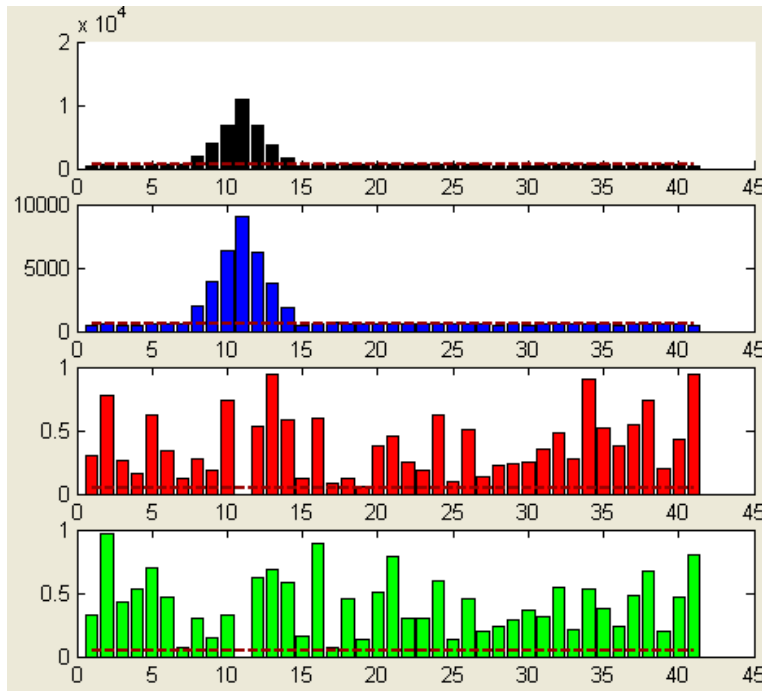


Ahora solo se ha modificado la dimensión del embedding, obteniéndose unos resultados similares al caso 1.

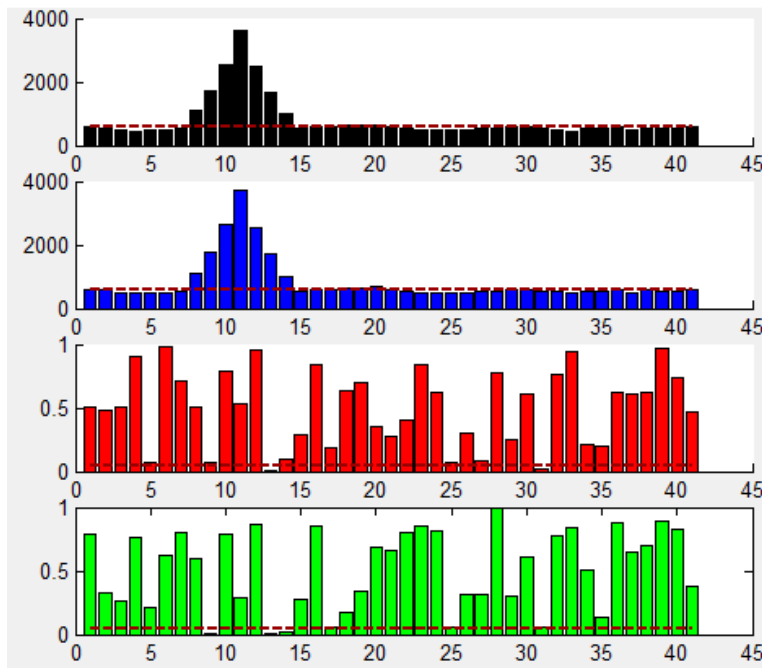
Caso 9: $m = 4$, $X(n) - Y_2(n)$, donde $Y_2(n) = 0'8 \cdot X(n - 1) + v(n)$



Caso 10: $m = 4$, $X(n) - Y_3(n)$, donde $Y_3(n) = 0'8 \cdot X(n - 10) + v(n)$



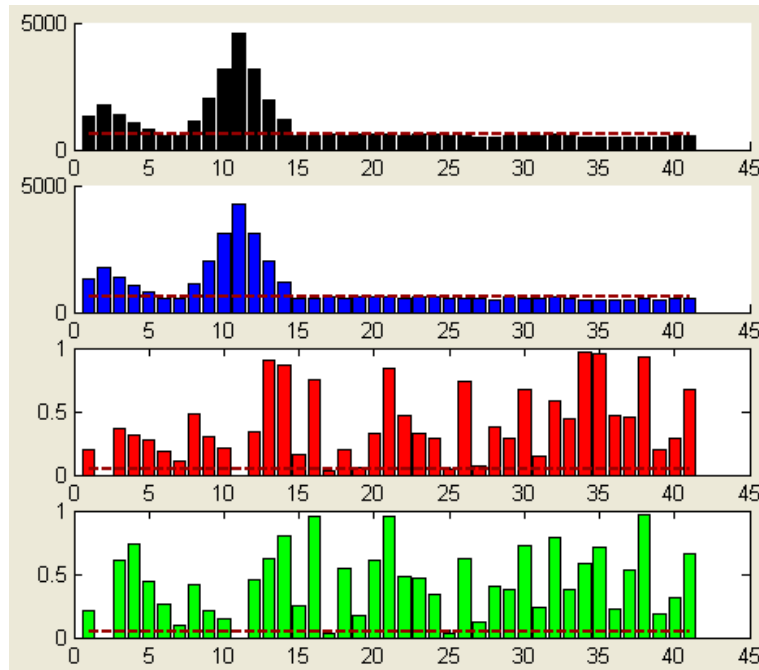
Caso 11: $m = 4$, $X(n) - Y_4(n)$, donde $Y_4(n) = 0'8 \cdot X(n - 10) \cdot X(n - 10) + v(n)$



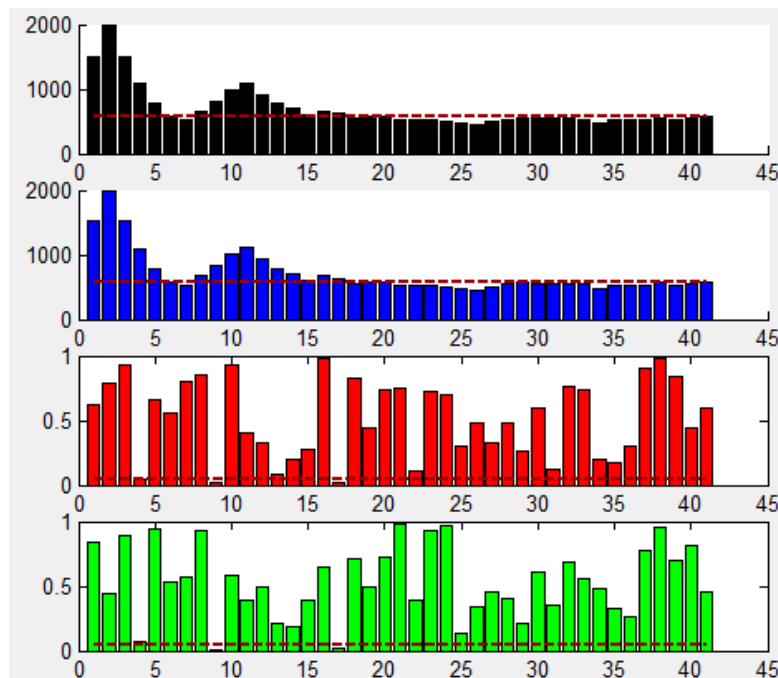
Igual caso que el 4, pero cambiando el embedding. A diferencia del caso 4, aquí los estadísticos de Pearson y Spearman no identifican de forma correcta donde se encuentra la máxima dependencia.



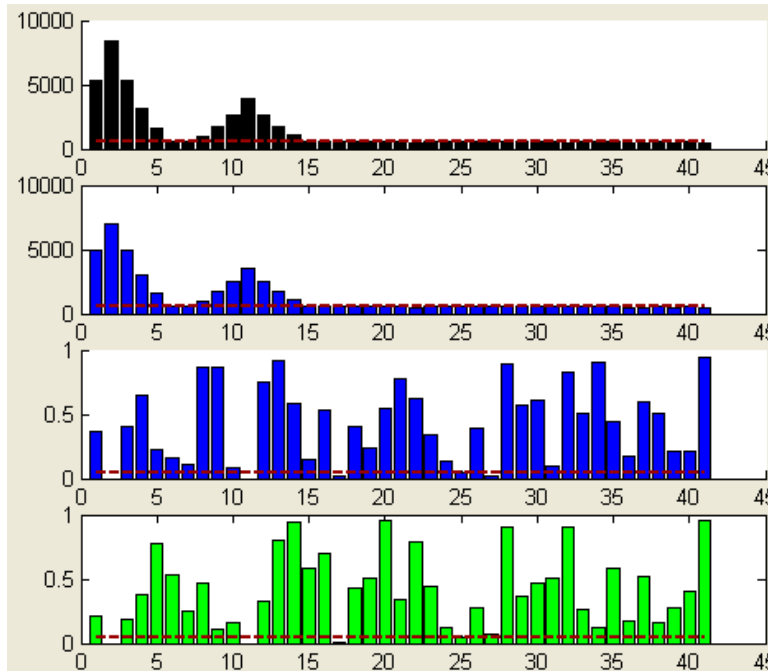
Caso 12: $m=4$, $X(n) - Y_5(n)$, donde $Y_5(n) = 0'3 \cdot X(n-1) + 0'5 \cdot X(n-10) + v(n)$



Caso 13: $m=4$, $X(n) - Y_6(n)$, donde $Y_6(n) = 0'5 \cdot X(n-1)^2 + 0'3 \cdot X(n-10)^2 + v(n)$



Caso 14: $m = 4$, $X(n) - Y_7(n)$, donde $Y_7(n) = 0'5 \cdot X(n-1)^3 + 0'3 \cdot X(n-10)^3 + v(n)$



Al aumentar la dimensión del embedding los resultados no cambian, aunque si se ve una pequeña diferencia significativa. Si para $m = 3$ se tiene que hay dependencia en los dos retardos contiguos al retardo de máxima dependencia, para $m = 4$ hay dependencia en tres retardos contiguos. Esto es normal ya que el estadístico ve dependencia en la forma de ordenar los términos dentro del mismo embedding.

En la segunda parte de este apartado, la serie $X(n)$ es un proceso AR(1) con coeficiente 0'9 para tener mayor grado de determinismo, mientras que la serie $Y(n)$ depende de la $X(n)$ de distintas formas (modelos $Y_1(n)$ hasta $Y_7(n)$), que son los utilizados en el apartado anterior:

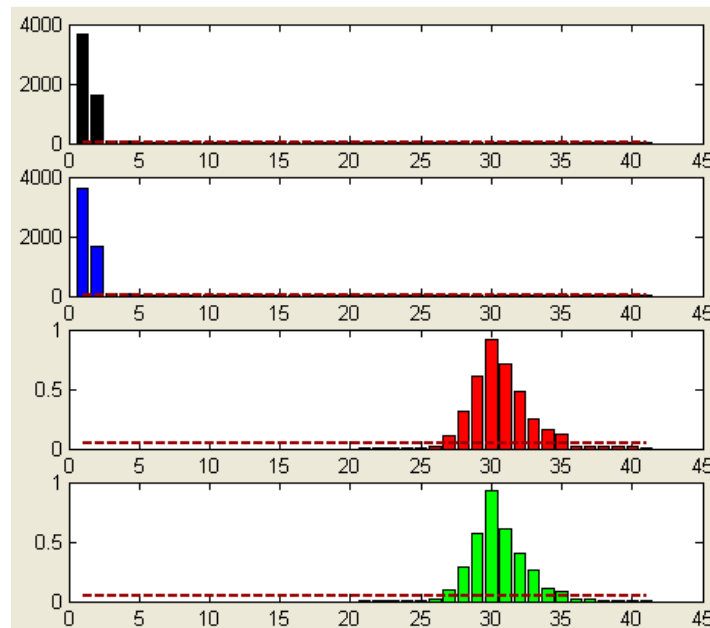
$$X(n) = 0'9 \cdot X(n-1) + u(n)$$

La longitud de cada una de las series es de 10000 datos. En este caso el tamaño de ventana será siempre el mismo y tendrá el mismo valor que la longitud de los datos. Lo que sí se modificará será la dimensión del embedding. Se realizarán las comparaciones para $m = 3$ y $m = 4$, de esta forma podremos ver el efecto de aumentar el embedding. Se mostrará un rango de retardos desde 0 hasta 40.

Para decidir si dos series son dependientes o independientes los criterios de aceptación y rechazo de la independencia son los mismos que en la primera parte de este apartado.

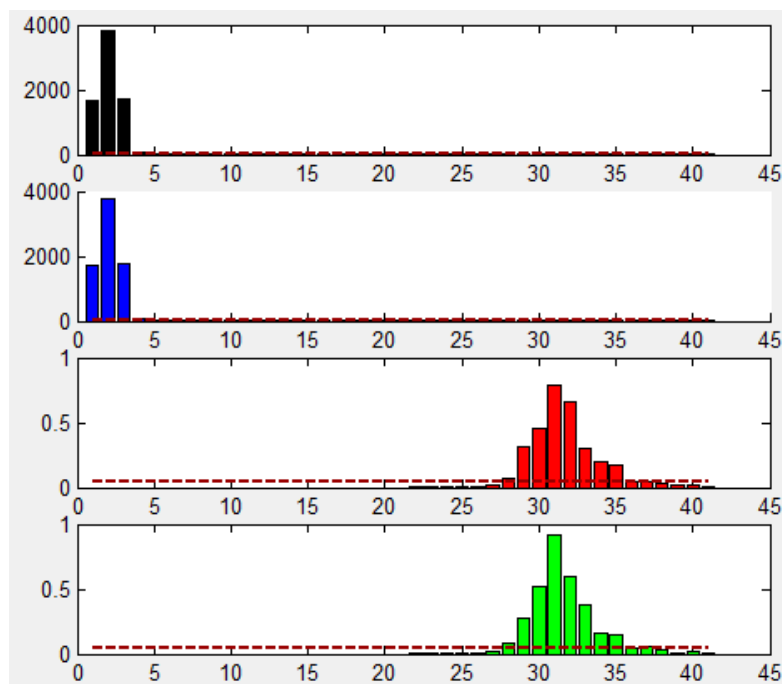
Se representarán los estadísticos para distintos retardos. La forma de representarlos también es la misma que en el anterior caso.

Caso 1: $m = 3$, $X(n) - Y_1(n)$, donde $Y_1(n) = 0'8 \cdot X(n) + v(n)$



Los estadísticos de Ji-cuadrado y G-test muestran una mayor dependencia con retardo 0 mientras que las correlaciones de Pearson y Spearman muestran una dependencia falsa en el retardo 29

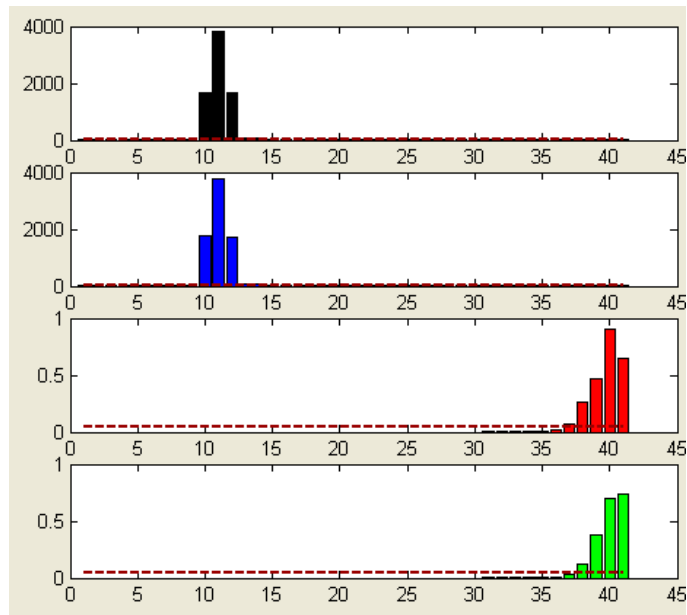
Caso 2: $m = 3$, $X(n) - Y_2(n)$, donde $Y_2(n) = 0'8 \cdot X(n - 1) + v(n)$



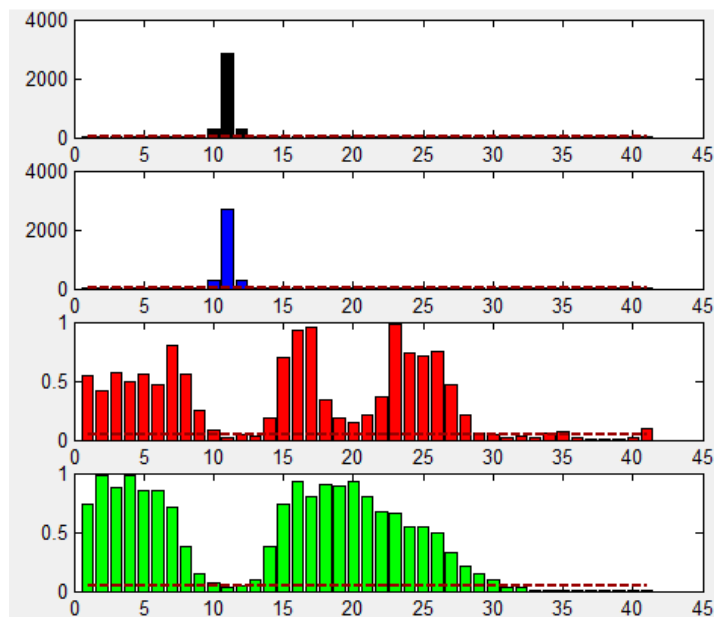
Al igual que en el caso anterior, solo detectan la dependencia de forma correcta Ji-cuadrado y G-test



Caso 3: $m = 3$, $X(n) - Y_3(n)$, donde $Y_3(n) = 0'8 \cdot X(n - 10) + v(n)$

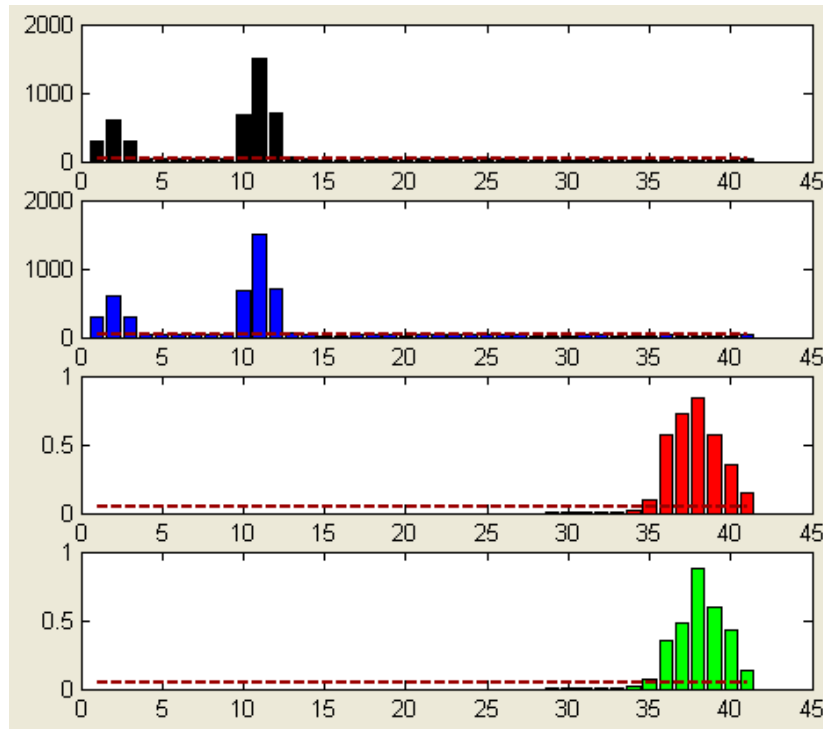


Caso 4: $m = 3$, $X(n) - Y_4(n)$, donde $Y_4(n) = 0'8 \cdot X(n - 10) \cdot X(n - 10) + v(n)$

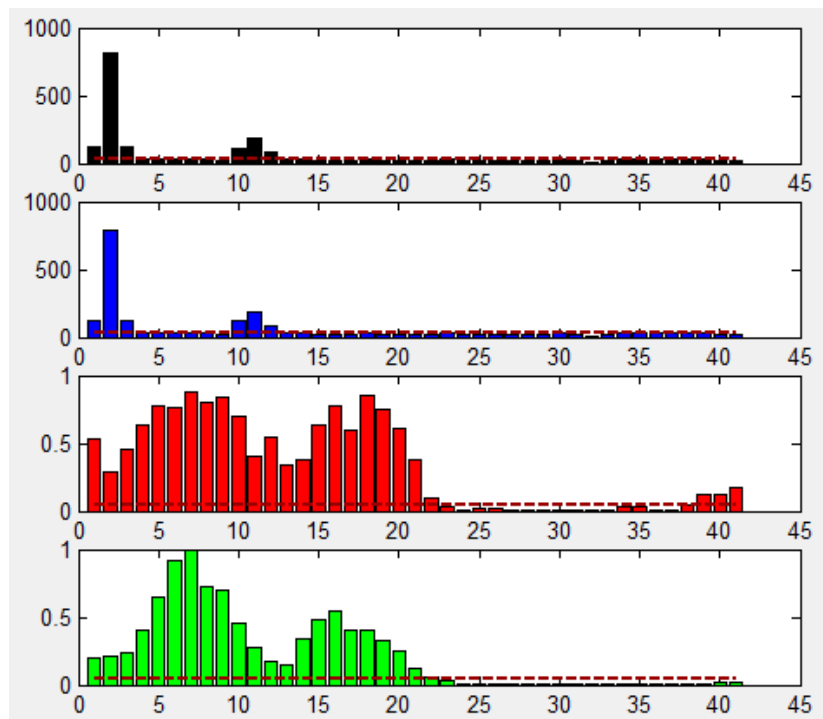




Caso 5: $m=3$, $X(n) - Y_5(n)$, donde $Y_5(n) = 0'3 \cdot X(n-1) + 0'5 \cdot X(n-10) + v(n)$



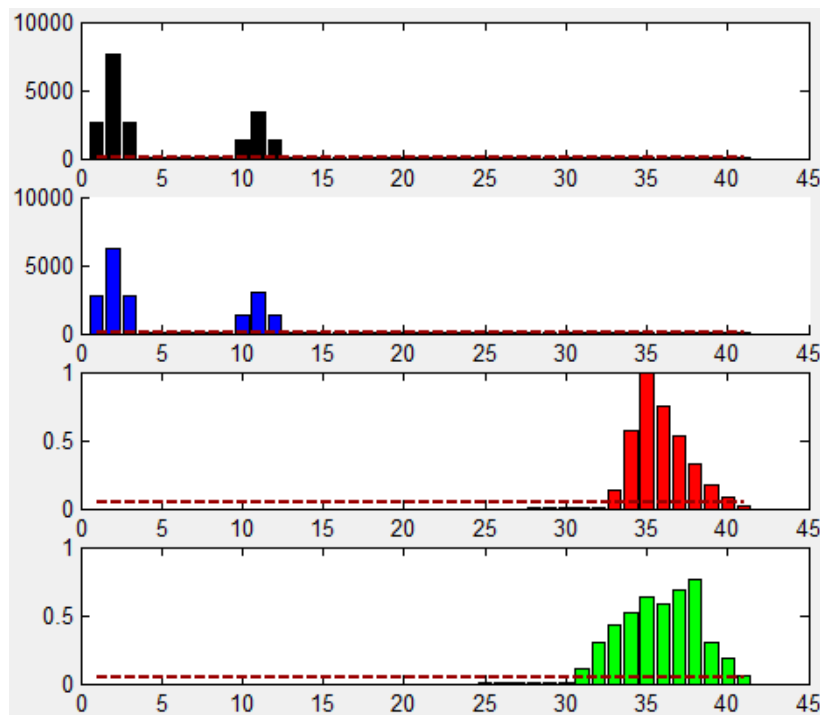
Caso 6: $m=3$, $X(n) - Y_6(n)$, donde $Y_6(n) = 0'5 \cdot X(n-1)^2 + 0'3 \cdot X(n-10)^2 + v(n)$



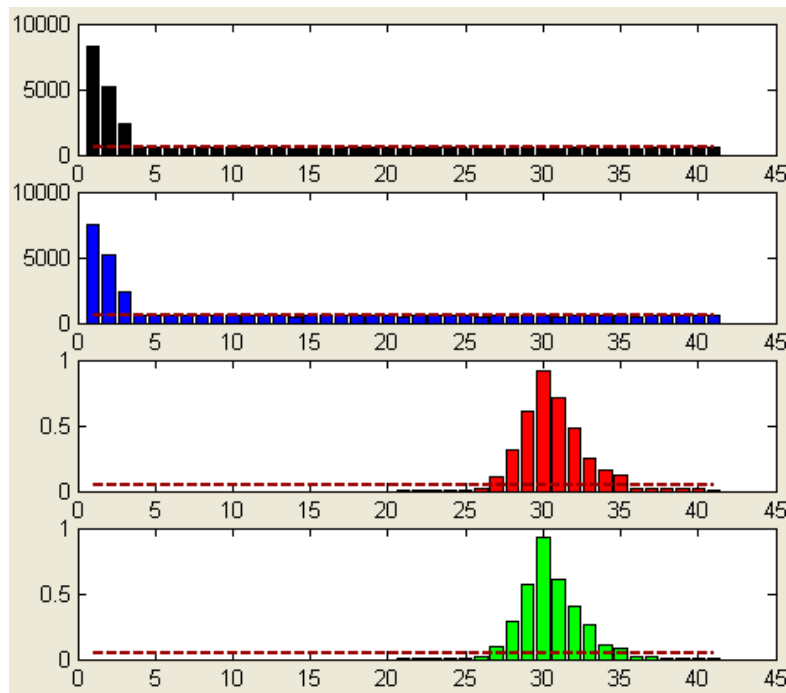


Caso 7: $m = 3$, $X(n) - Y_7(n)$, donde $Y_7(n) = 0'5 \cdot X(n-1)^3 + 0'3 \cdot X(n-10)^3 +$

$$v(n)$$

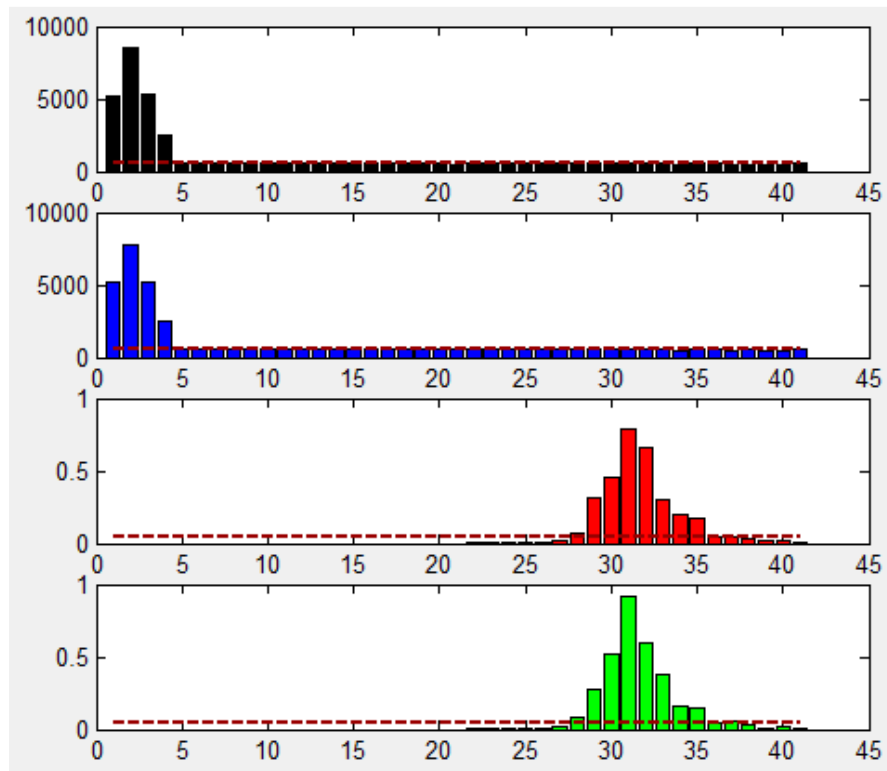


Caso 8: $m = 4$, $X(n) - Y_1(n)$, donde $Y_1(n) = 0'8 \cdot X(n) + v(n)$

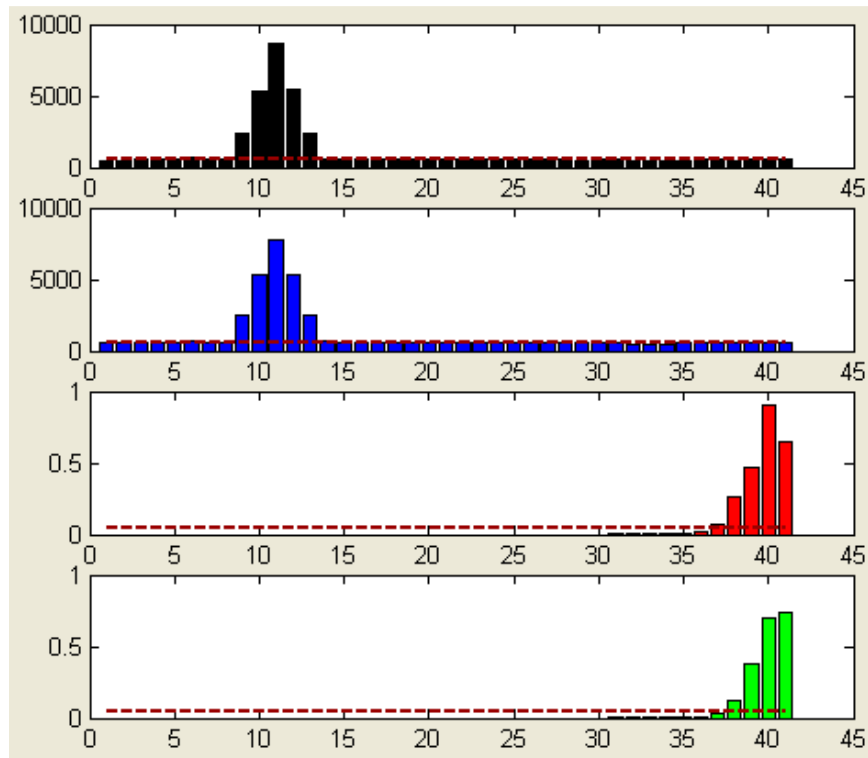




Caso 9: $m = 4$, $X(n) - Y_2(n)$, donde $Y_2(n) = 0'8 \cdot X(n - 1) + v(n)$

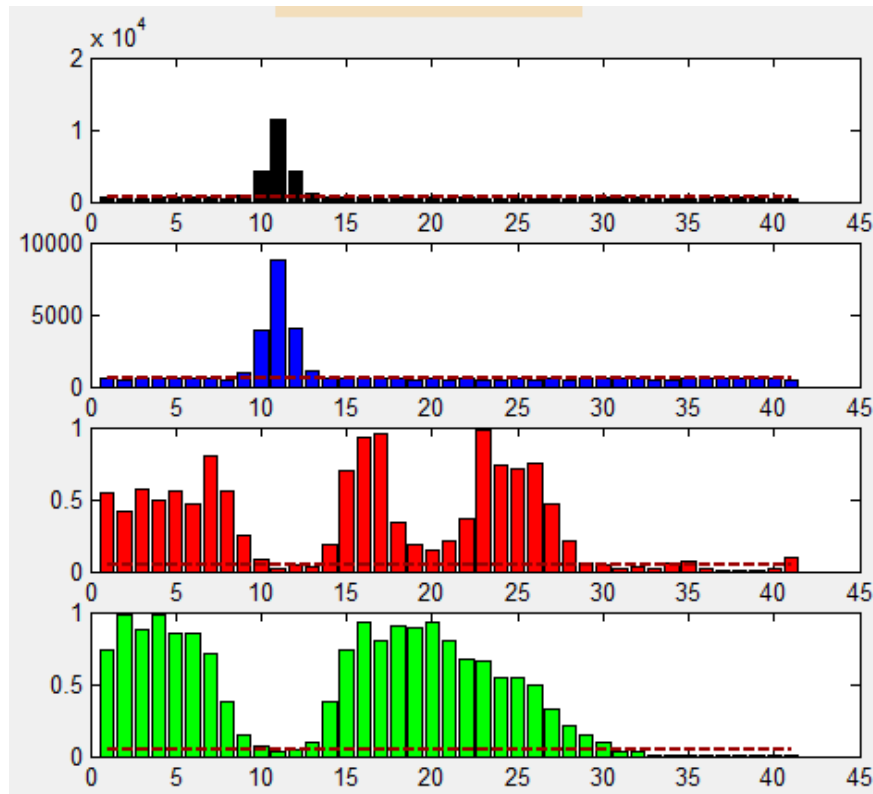


Caso 10: $m = 4$, $X(n) - Y_3(n)$, donde $Y_3(n) = 0'8 \cdot X(n - 10) + v(n)$

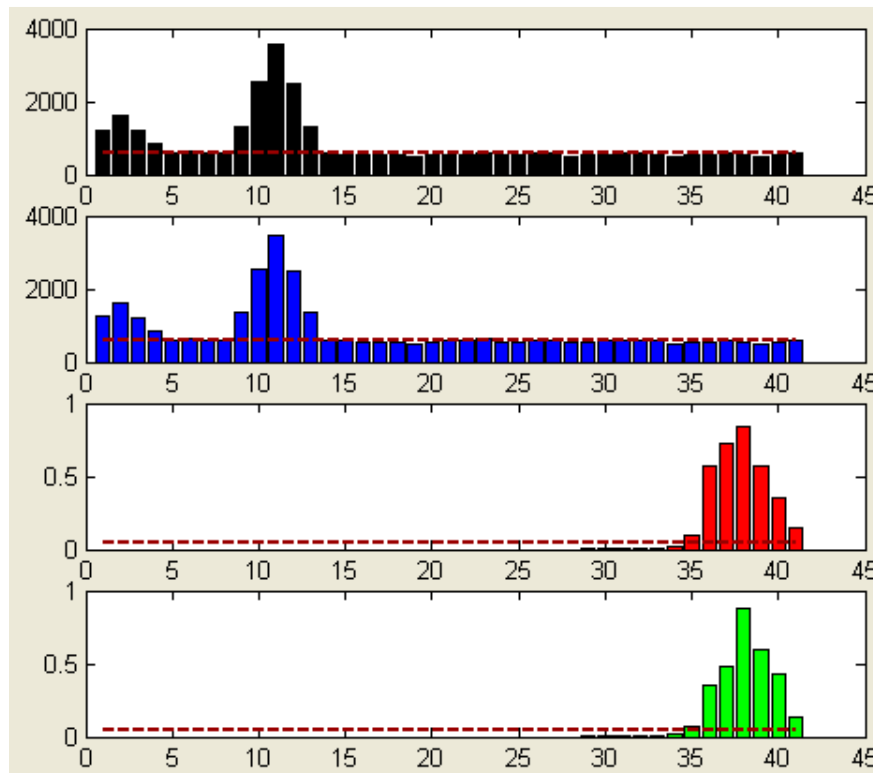




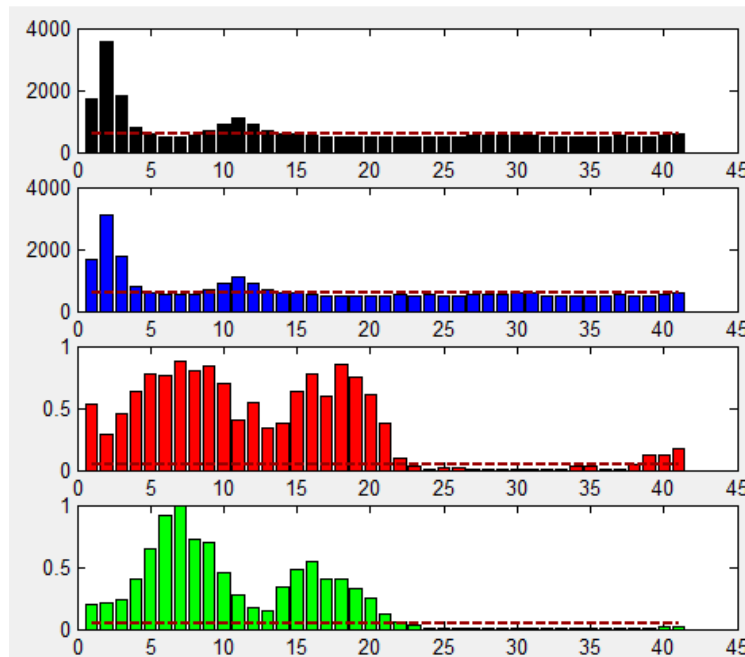
Caso 11: $m = 4$, $X(n) - Y_4(n)$, donde $Y_4(n) = 0'8 \cdot X(n - 10) \cdot X(n - 10) + v(n)$



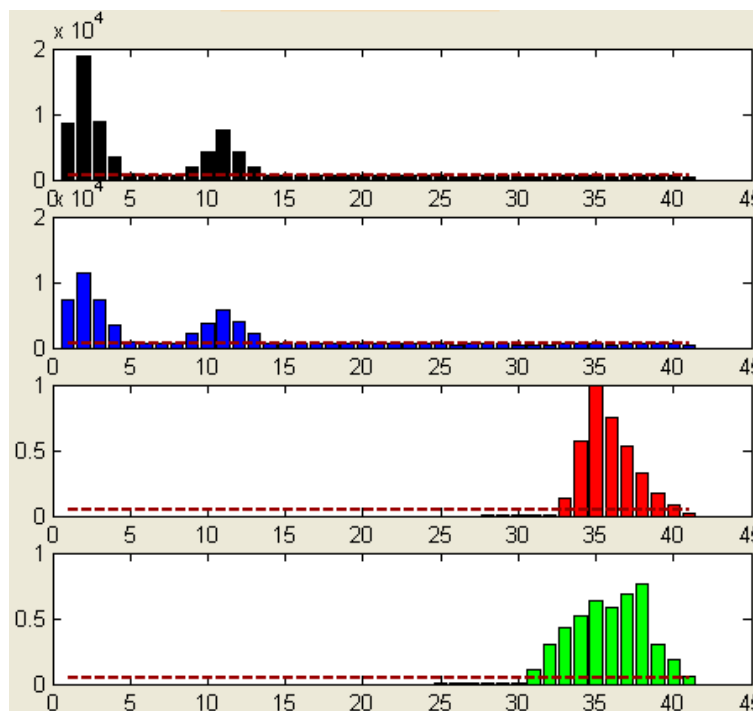
Caso 12: $m=4$, $X(n) - Y_5(n)$, donde $Y_5(n) = 0'3 \cdot X(n - 1) + 0'5 \cdot X(n - 10) + v(n)$



Caso 13: $m=4$, $X(n) - Y_6(n)$, donde $Y_6(n) = 0'5 \cdot X(n-1)^2 + 0'3 \cdot X(n-10)^2 + v(n)$



Caso 14: $m = 4$, $X(n) - Y_7(n)$, donde $Y_7(n) = 0'5 \cdot X(n-1)^3 + 0'3 \cdot X(n-10)^3 + v(n)$



Cuando usamos el correlograma de Pearson y Spearman (o p-valores), vemos que para las dependencias lineales hay fallos.



5. APLICACIÓN A LOS MERCADOS ELÉCTRICOS

En este capítulo aplicaremos la metodología propuesta (cálculo de medidas de asociación y de distancia por ventanas móviles) a series de precios de la electricidad de distintos mercados. Distinguiremos 2 secciones: la primera está dedicada al estudio de autodependencias en los mercados eléctricos, mientras que la segunda está dedicada al estudio de dependencias entre distintos mercados eléctricos.

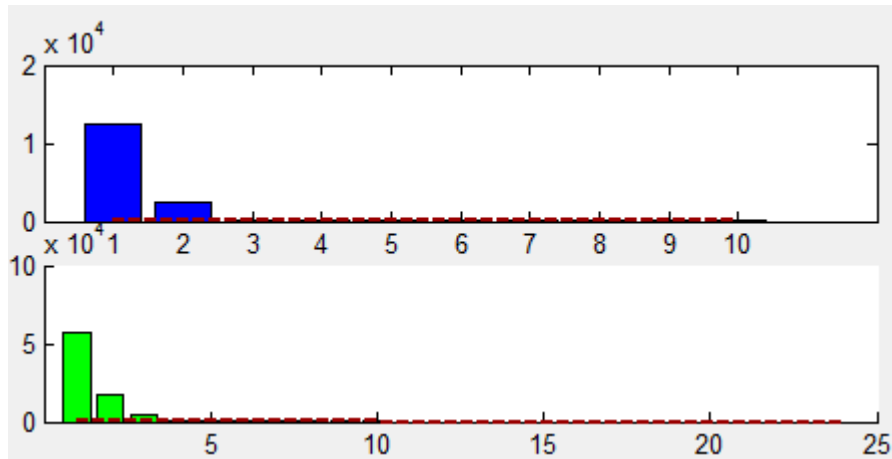
Para nuestro análisis hemos considerado los siguiente Mercados Eléctricos: OMEL (mercado español), Austria, Nordpool (sistema, Suecia, Finlandia, Noruega, Dinamarca Este y Dinamarca Oeste), Ontario y AEMO (las cuatro regiones de Australia: Victoria, Queensland, Nueva Gales del Sur, Australia del Sur). Los motivos para seleccionar estos mercados han sido dos principalmente. Por un lado, se trata de Mercados Eléctricos de relevancia a nivel europeo e internacional, y por otro lado, sus datos eran de libre acceso. Con respecto a esto último, cabe destacar que una parte importante del proyecto ha sido la recopilación, restructuración en formato adecuado y sustitución de valores perdidos de los distintos conjuntos de datos. Además, con el fin de disponer de un rango común de tiempo, se optó por seleccionar datos históricos (precios de la electricidad por horas) desde Enero 2004 hasta Diciembre 2009, para cada uno de los Mercados mencionados anteriormente.

5.1. AUTODEPENDENCIAS EN MERCADOS ELÉCTRICOS

En esta sección se incluirán los resultados de estudiar autodependencias, es decir, asociación de una serie de precios consigo misma retardada en el tiempo. Lo que se pretende es determinar qué instantes en el pasado pueden producir una influencia significativa sobre el precio actual. Se hará especial énfasis en los retardos $h=1$, $h=24$, $h=168$ correspondientes a una hora, un día y una semana.

En primer lugar, destacamos un “inconveniente” del método propuesto: cualquier serie (incluyendo un ruido o perturbación aleatoria), va a presentar dependencia con su retardo de orden 1, de orden 2, ... hasta orden $(m-1)$, siendo $m=\text{embedding}$. Esta situación ocurre por la naturaleza del método propuesto.

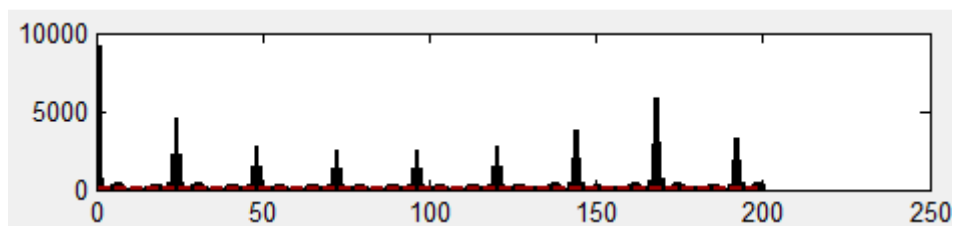
Ejemplo ilustrativo: Consideremos una serie de longitud 10000 datos procedente de un ruido blanco o perturbación aleatoria. Vamos a analizar la autodependencia existente para el rango de retardos desde 1 hasta 10, utilizando el estadístico Chi-cuadrado. Primero lo haremos con embedding $m = 3$ y luego con un $m = 4$.



El primero de los gráficos se corresponde a $m = 3$, y el segundo a $m = 4$. Los resultados deberían mostrar que no existe dependencia para ningún retardo por tratarse de un ruido, pero como puede verse, hay dependencia significativa con su retardo hasta orden $m-1$.

Una vez indicada la limitación anterior, pasamos a realizar el análisis de autodependencias para cada mercado. Analizaremos los mercados OMEL, Austria, Ontario, Nordpool y Australia (Victoria). Se usará primero el programa “retardo” para comprobar que los retardos de máxima dependencia son $h=1$, $h=24$, $h=168$, con $m = 3$, usando los datos del primer año (solo se mostrará el rango de retardos desde 1 a 200). A continuación se usará el programa Dependencias con un retardo fijo para estudiar si el nivel de dependencia es constante a lo largo del tiempo o no. Se ha seleccionado un tamaño de ventana $w=2190$, que corresponde a la duración de una estación, así se podrá analizar si la dependencia es más o menos constante a lo largo las estaciones

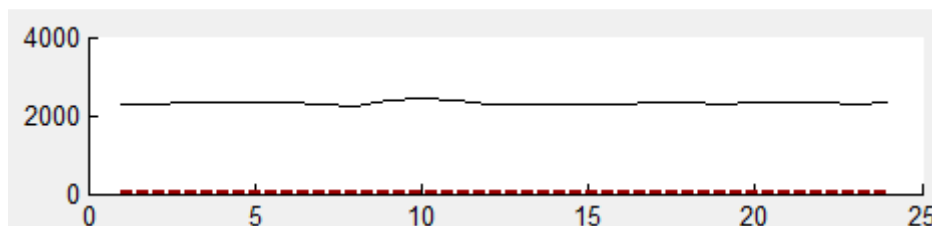
OMEL



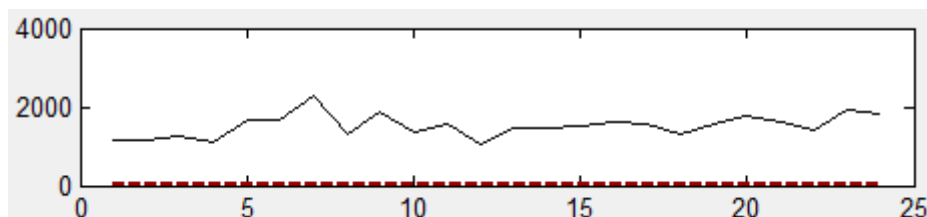
Se observa que las mayores dependencias están en los retardos 1, 24 (0 múltiplos de 24) y 168, siendo algo mayor la dependencia semanal que la diaria.

A continuación mostramos los resultados de poder aplicar el programa Dependencias para los retardos $h=1$, 24 y 168, para estudiar si el nivel de dependencia es constante a lo largo del tiempo

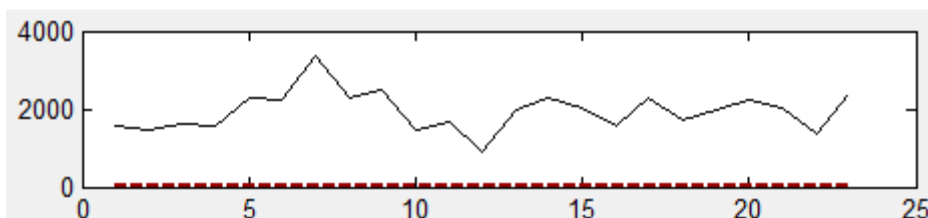
- $h=1$



- $h=24$

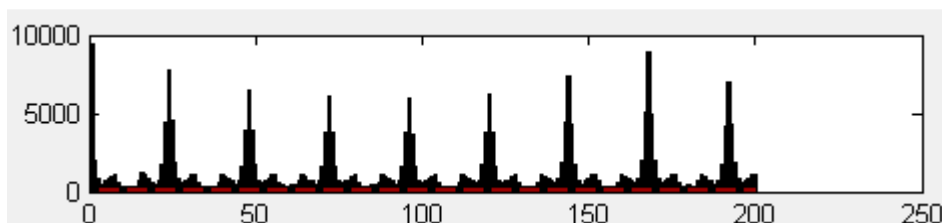


- $h=168$



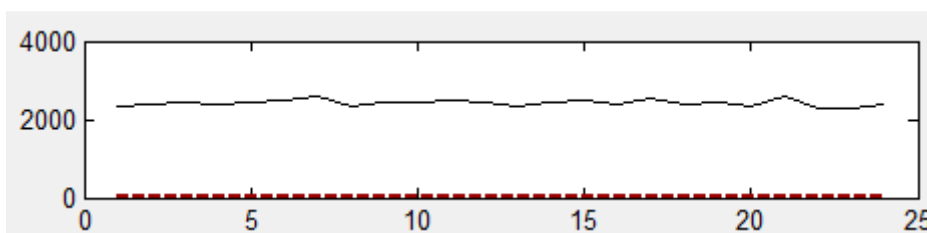
Para $h=1$ resulta una medida de dependencia muy homogénea a lo largo de las estaciones. Para el retardo $h=24$ y $h=168$ se observan picos de dependencia en los veranos, luego parece que la estacionalidad diaria y semanal es más acentuada en los veranos que en el resto de estaciones, destacando una fuerte dependencia en verano de 2005 tanto diaria como semanal.

Austria

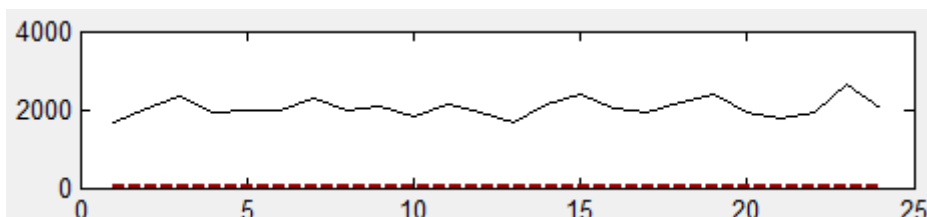


Al igual que en el mercado OMEL, en Austria las mayores dependencias están en los retardos 1, 24 y múltiplos de 24, siendo ligeramente superior la semanal que la diaria. Además, podemos comprobar que dichas componentes estacionales son más fuertes que en el mercado español. Veamos si el nivel de asociación es homogénea a lo largo de los 6 años de estudio.

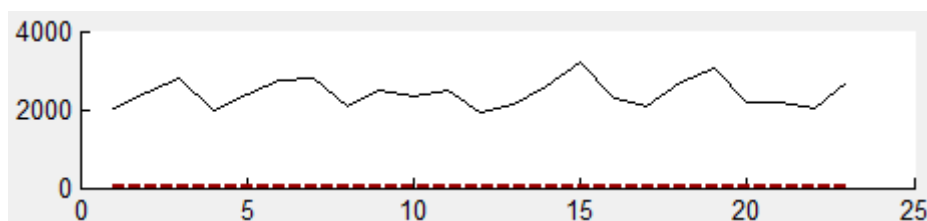
- $h=1$



- $h=24$



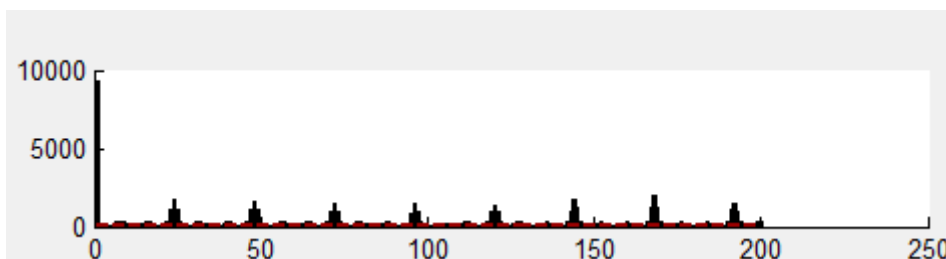
- $h=168$



En el caso de Austria ocurre lo mismo que en OMEL: para $h=1$ hay dependencia muy homogénea a lo largo de las estaciones, mientras que para $h=24$ y $h=168$ existen picos de dependencia en los veranos.

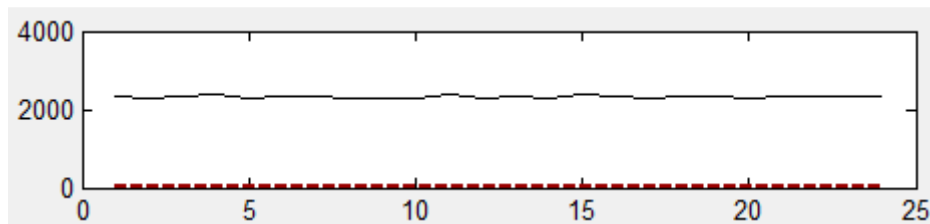
Por tanto, en estos mercados la componente estacional diaria y semanal es más acentuada en los veranos que en el resto de estaciones, lo que permitiría mejores modelos de previsión de precios.

Ontario

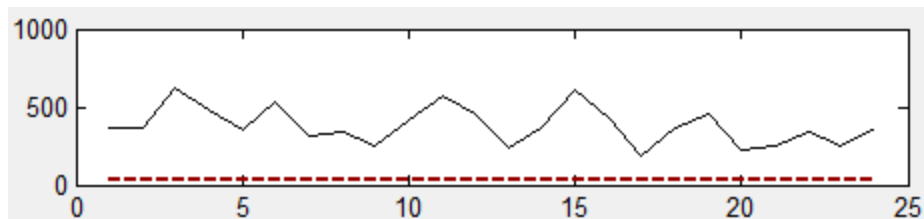


Como sucedía con los otros mercados, las máximas dependencias siguen siendo para los retardos 1, 24 y múltiplos de 24, siendo ligeramente superior la dependencia semanal frente a la diaria. Conviene destacar que para Ontario las dependencias estacionales son mucho menores que las que eran para OMEL y Austria. Veamos si el grado de asociación es variable a lo largo de los 6 años en estudio.

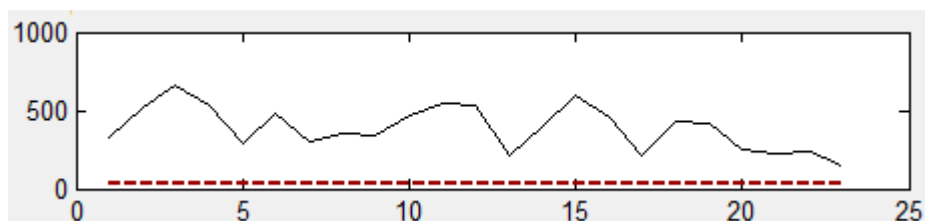
- $h=1$



- $h=24$

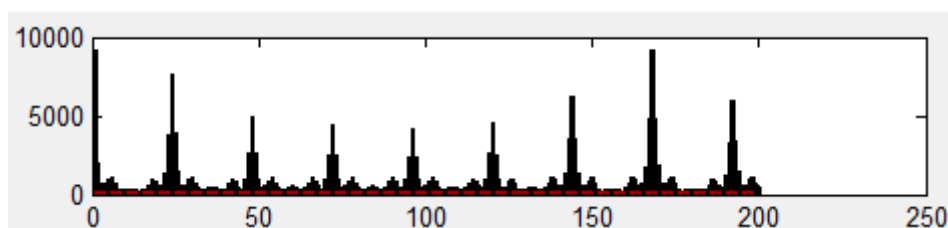


- $h=168$



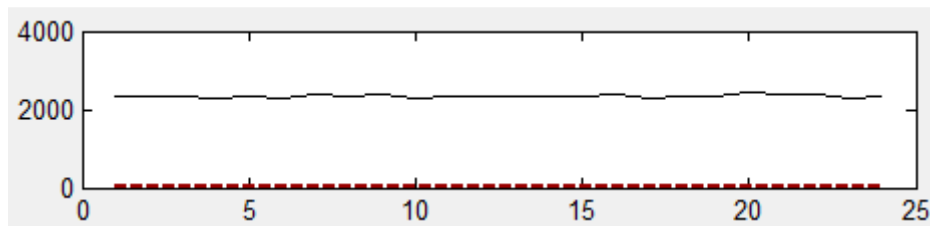
En Ontario, sucede algo parecido a OMEL y Austria, es decir, las dependencias más fuertes suelen darse en los veranos, salvo en algún año puntual.

Nordpool

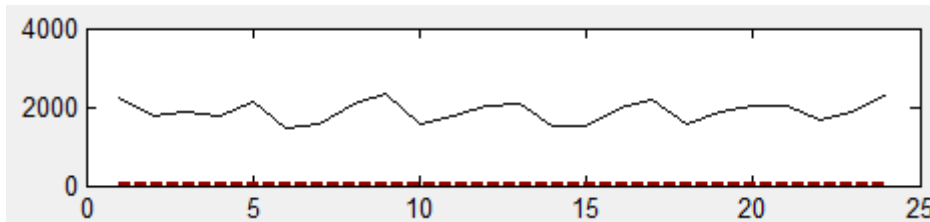


Nuevamente, las mayores asociaciones se producen para los retardos 1, 24 y múltiplos de 24, destacando principalmente la dependencia semanal ($h=168$). Además, el nivel o grado de asociación estacional es del estilo de Austria y por tanto ligeramente superior a OMEL y mucho superior a Ontario. Veamos cómo varía el grado de asociación a lo largo de los años en estudio

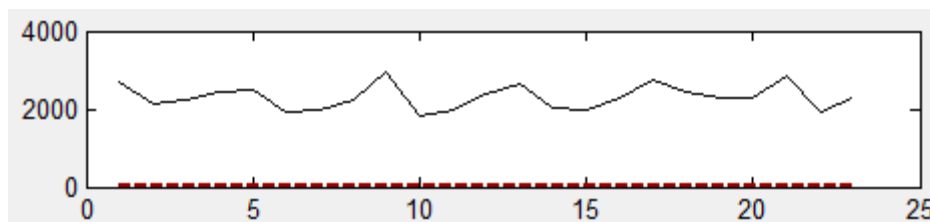
- $h=1$



- $h=24$

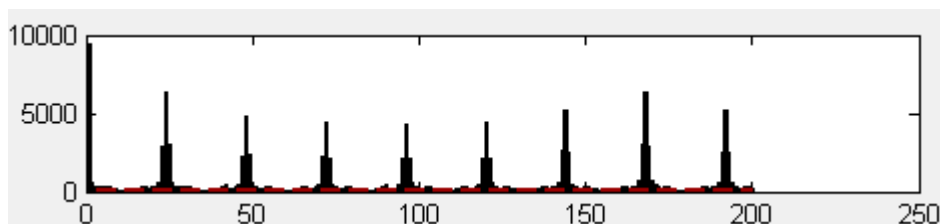


- $h=168$



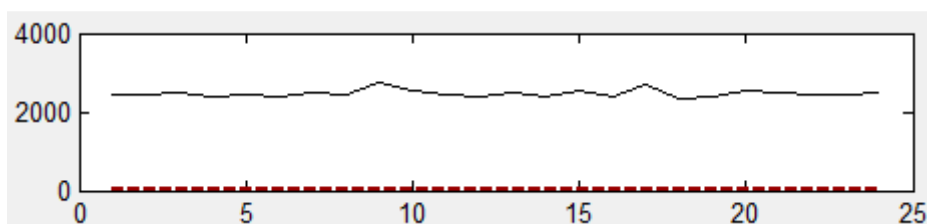
En Nordpool la dependencia vuelve a ser muy homogénea a lo largo de los años para el retardo en $h=1$. Sin embargo, y a diferencia de los otros mercados, la estacionalidad diaria y semanal es más acentuada en los inviernos que en el resto de estaciones.

Australia (Victoria)

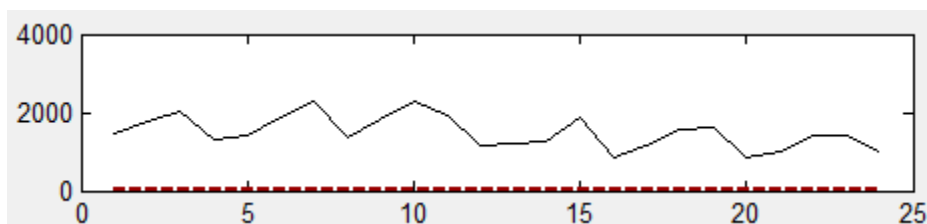


Una vez más el mismo esquema: máximas dependencias en los retardos 1, 24 y múltiplo de 24, con niveles de asociación algo inferiores a Nordpool.

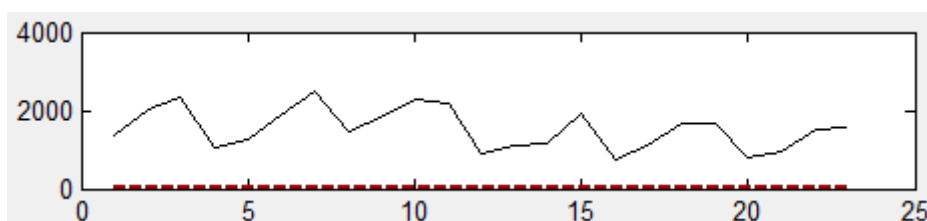
- $h=1$



- $h=24$



- $h=168$



En Australia, la dependencia diaria y semanal es más acentuada en los inviernos (verano europeo), aunque en 2006 tuvo su máximo en otoño.

5.2. DEPENDENCIAS ENTRE PRECIOS DE DISTINTOS MERCADOS

En esta sección se estudian dependencias entre series de precios de distintos mercados, vecinos y no vecinos (con y sin proximidad geográfica), y regulados o no por el mismo operador (casos de AEMO con 4 zonas y NordPool con varias zonas).

Se comentará qué mercados presentan mayores y menores dependencias con las series sincronizadas en el tiempo, y también con las series usando retardos. También se destacará en qué épocas del año se dan las relaciones más y menos estrechas.

Se utilizarán los datos de precios horarios de cada mercado referidos al intervalo 2004-2009, ambos inclusive. Teniendo en cuenta la influencia de la componente estacional en el análisis de dependencias, véase Capítulo 4, sección 2, en primer lugar extraeremos a cada Mercado su componente estocástica (eliminamos la componente estacional), de 3 formas distintas:



- Tomando diferencias estacionales semanales.
- Tomando diferencias estacionales semanales y luego diarias.
- Realizando una ponderación de diferencias semanales y diarias (véase Weron (2006)).

Ahora, vamos a concentrarnos en los métodos de descomposición de la señal.

La diferenciación es una técnica que se puede utilizar para eliminar los componentes estacionales y las tendencias. Hay dos razones principales para la diferenciación de una serie. En primer lugar, se puede identificar la naturaleza oculta de las dependencias temporales de las series. Quitando algunas de las autocorrelaciones cambiaremos otras autocorrelaciones, ya que son interdependientes para retardos consecutivos. La diferenciación puede eliminarlas o hacer algunas otras estacionalidades más evidentes. La otra razón para eliminar las dependencias temporales es hacer las series estacionarias, lo cual es necesario para el modelado de series temporales.

La idea que hay detrás de la diferenciación no es más que tener en cuenta las diferencias entre los sucesivos pares de observaciones con los retardos apropiados. Por ejemplo, una tendencia lineal puede ser eliminada mediante la diferenciación en retardo 1, mientras que una tendencia de un polinomio de orden k -ésimo por diferenciación de k veces en el retardo 1. Del mismo modo, para eliminar una componente estacional semanal de datos diarios (x_1, \dots, x_n), se genera la serie transformada:

$$y_t = x_t - x_{t-7} = (1 - B^7)x_t$$

Donde B es el operador de retardos o de desplazamiento hacia atrás, es decir $B^h x_t \equiv x_{t-h}$. Aparentemente toda componente estacional de período 7 es eliminada por esta transformación. Tenemos que ser cautelosos al interpretar los correlogramas, ya que algunas de las dependencias observadas pueden ser falsas debido al proceso de diferenciación

Cuando se trate de datos horarios debemos tener en cuenta dos estacionalidades: semanal y diaria. Lo más simple es aplicar un retardo 168, porque el período diario ya está incluido por la estacionalidad semanal. Sin embargo, un enfoque más popular entre los profesionales es utilizar diferenciación en diversos retardos (por lo general 1, 24 y 168), a menudo en combinación con la media móvil de tipo suavizado. Por ejemplo, el componente estocástico podría ser extraído de los datos horarios como (véase Weron (2006)):

$$y_t = x_t - \left(\frac{1}{N} \sum_{i=1}^N x_{t-i \cdot 168} + \frac{1}{7} \sum_{j=1}^7 x_{t-j \cdot 24} - \frac{1}{7N} \sum_{i=1}^N \sum_{j=1}^7 x_{t-i \cdot 168 - j \cdot 24} \right)$$

Donde $N+1 = 5$ o 6 es el número de semanas utilizadas para la calibración.

A modo de ejemplo, estudiaremos la dependencia entre OMEL y Ontario en 4 situaciones distintas:

- (1) Usando datos originales, es decir, sin eliminar la componente estacional
- (2) Tomando diferencias estacionales semanales en los datos.
- (3) Tomando diferencias estacionales semanales en los datos.
- (4) Realizando la desestacionalización propuesta en Weron (2006).

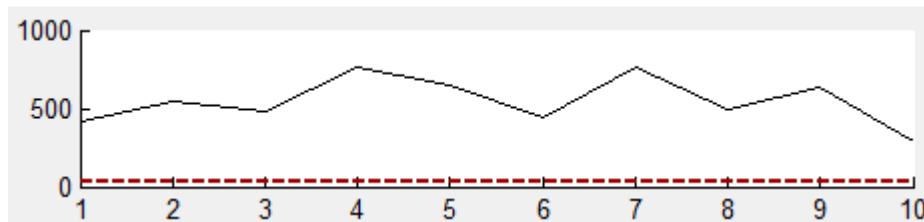


Figura 5.1: Análisis de dependencias para datos originales caso (1).

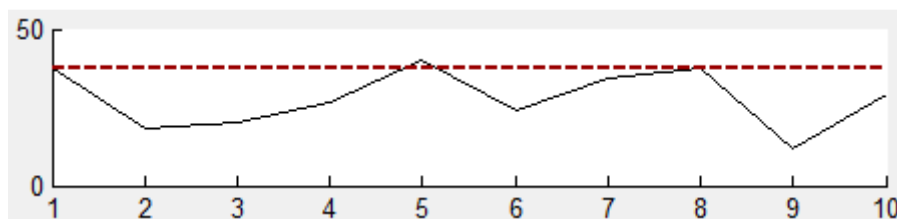


Figura 5.2: Análisis de dependencias para datos desestacionalizados caso (2).

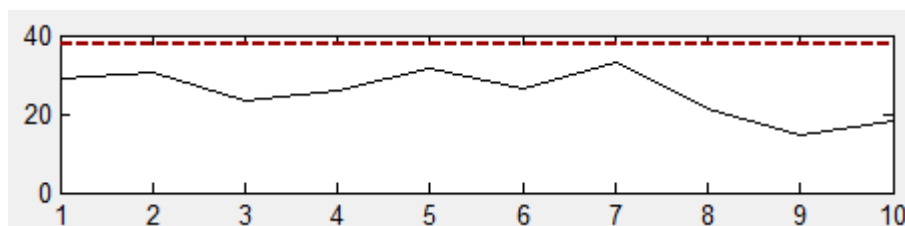


Figura 5.3: Análisis de dependencias para datos desestacionalizados caso (3).

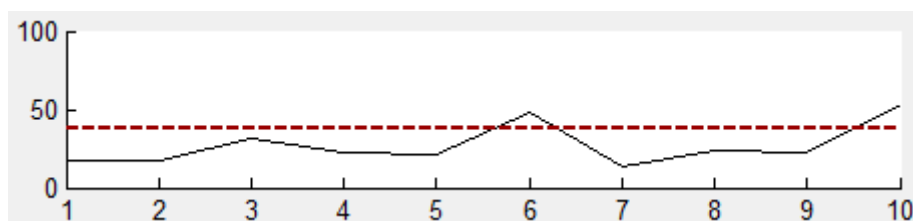


Figura 5.4: Análisis de dependencias para datos desestacionalizados caso (4).

Destacar que en el caso (1) sale una falsa dependencia debida a la fuerte componente estacional diaria y semanal presente en ambas series, mientras que en el resto de casos sale independencia entre ambos mercados, lo que se ajusta a la realidad. Para esta parte se han las medidas Chi-cuadrado, G-test y Cressi Read, para disponer de límites de aceptación y rechazo de la independencia, aunque solo se muestra Chi-cuadrado, ya que los resultados son similares para los tres estadísticos. Se ha usado $m=3$ y $w=5000$.

Ahora, con los datos tipo Weron y usando los dos mercados más estrechamente relacionados (Finlandia y Suecia), con $m=3$ y $w=2190$ (una estación), analizaremos si la dependencia es homogénea a lo largo de los 6 años o si es variable. Para ello, se ha representado el estadístico Ji-cuadrado porque el resto son similares.

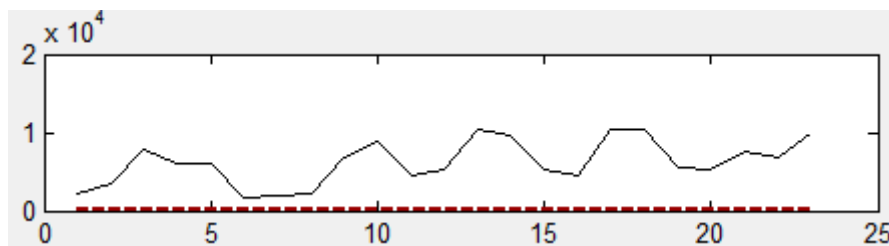
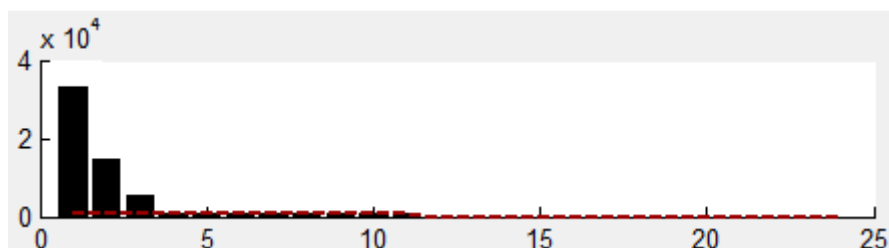
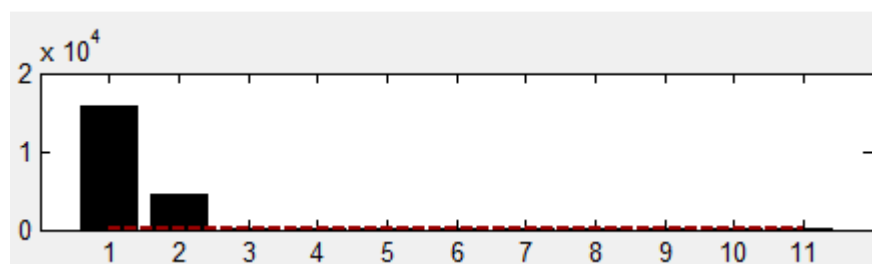


Figura 5.5: Análisis de dependencias entre Finlandia y Austria

Se observa que la dependencia no es homogénea a lo largo del periodo en estudio. Existen picos (máximos relativos) en otoño de 2004, verano 2006, primavera-verano 2007, primavera-verano 2008 y en otoño 2009. Además, se aprecia un ligero aumento en el nivel de asociación a lo largo de los años.

Para estos mismos mercados vamos a analizar cuál sería el retardo en el tiempo que proporciona máxima dependencia. En el siguiente gráfico mostramos el estadístico Ji-cuadrado para el rango de retardos desde 0 hasta 10, con $m = 3$ y $m = 4$.





Se observa que el retardo de máxima dependencia es $h=0$, es decir, los precios de ambos mercados presentan relación más estrecha en el caso de las series sincronizadas en el tiempo.

Para finalizar el capítulo, vamos a analizar cuál sería la estructura de dependencia entre todos los mercados en estudio, calculando para ello la matriz de distancias entre pares de mercados. Posteriormente, obtendremos el árbol de expansión mínima y la clasificación de los mercados en clusters homogéneos.

En primer lugar, cada medida de “distancia” (distancia V de Cramer, distancia del Coeficiente de Incertidumbre, distancia Universal 1 y distancia Universal 2) se calculará la Matriz de Distancias entre cada par de Mercados. Para ello se usará los datos del tipo Weron. Como intervalo de tiempo, consideraremos primero el rango completo de 2004-2009 como ventana única, y más tarde calcularemos la matriz de distancias para cada año completo (2005, 2006, 2007, 2008 y 2009). Se ha seleccionado $m=3$ para ventanas anuales y $m=4$ para la ventana única de 6 años.

Global (2004-2009)

<u>V Cramer</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,65	0,77	0,65	0,98	0,98	0,98	0,98	0,97	0,97	0,98	0,98	0,98	0,98
QLD	0,65	0,00	0,86	0,80	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98
SA	0,77	0,86	0,00	0,66	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98
VIC	0,65	0,80	0,66	0,00	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98
Austria	0,98	0,98	0,98	0,98	0,00	0,96	0,95	0,97	0,97	0,97	0,98	0,97	0,97	0,97
Din. E	0,98	0,98	0,98	0,98	0,96	0,00	0,89	0,75	0,85	0,98	0,98	0,92	0,68	0,86
Din. O	0,98	0,98	0,98	0,98	0,95	0,89	0,00	0,92	0,91	0,98	0,98	0,95	0,91	0,94
Finlandia	0,98	0,98	0,98	0,98	0,97	0,75	0,92	0,00	0,79	0,98	0,98	0,90	0,35	0,74
Nordpool	0,97	0,98	0,98	0,98	0,97	0,85	0,91	0,79	0,00	0,98	0,98	0,76	0,75	0,79
OMEL	0,97	0,98	0,98	0,98	0,97	0,98	0,98	0,98	0,98	0,00	0,98	0,98	0,98	0,98
Ontario	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,00	0,98	0,98	0,98
Oslo	0,98	0,98	0,98	0,98	0,97	0,92	0,95	0,90	0,76	0,98	0,98	0,00	0,86	0,87
Suecia	0,98	0,98	0,98	0,98	0,97	0,68	0,91	0,35	0,75	0,98	0,98	0,86	0,00	0,63
Tro	0,98	0,98	0,98	0,98	0,97	0,86	0,94	0,74	0,79	0,98	0,98	0,87	0,63	0,00



<u>Coef de Incertid.</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,75	0,87	0,75	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,75	0,00	0,94	0,89	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,87	0,94	0,00	0,76	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,75	0,89	0,76	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	0,99	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	0,99	0,00	0,96	0,87	0,93	1,00	1,00	0,98	0,82	0,94
Din. O	1,00	1,00	1,00	1,00	0,99	0,96	0,00	0,98	0,97	1,00	1,00	0,99	0,97	0,99
Finlandia	1,00	1,00	1,00	1,00	1,00	0,87	0,98	0,00	0,87	1,00	1,00	0,97	0,48	0,85
Nordpool	1,00	1,00	1,00	1,00	1,00	0,93	0,97	0,87	0,00	1,00	1,00	0,85	0,82	0,87
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,98	0,99	0,97	0,85	1,00	1,00	0,00	0,94	0,95
Suecia	1,00	1,00	1,00	1,00	1,00	0,82	0,97	0,48	0,82	1,00	1,00	0,94	0,00	0,76
Tro	1,00	1,00	1,00	1,00	1,00	0,94	0,99	0,85	0,87	1,00	1,00	0,95	0,76	0,00

<u>Universal</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,86	0,93	0,86	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,86	0,00	0,97	0,94	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,93	0,97	0,00	0,86	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,86	0,94	0,86	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	1,00	0,00	0,98	0,93	0,96	1,00	1,00	0,99	0,90	0,97
Din. O	1,00	1,00	1,00	1,00	1,00	0,98	0,00	0,99	0,98	1,00	1,00	1,00	0,99	0,99
Finlandia	1,00	1,00	1,00	1,00	1,00	0,93	0,99	0,00	0,93	1,00	1,00	0,98	0,65	0,92
Nordpool	1,00	1,00	1,00	1,00	1,00	0,96	0,98	0,93	0,00	1,00	1,00	0,92	0,90	0,93
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,99	1,00	0,98	0,92	1,00	1,00	0,00	0,97	0,97
Suecia	1,00	1,00	1,00	1,00	1,00	0,90	0,99	0,65	0,90	1,00	1,00	0,97	0,00	0,86
Tro	1,00	1,00	1,00	1,00	1,00	0,97	0,99	0,92	0,93	1,00	1,00	0,97	0,86	0,00



<u>Universal</u> <u>2</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,75	0,87	0,75	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,75	0,00	0,94	0,89	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,87	0,94	0,00	0,76	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,75	0,89	0,76	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	0,99	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	0,99	0,00	0,96	0,87	0,93	1,00	1,00	0,98	0,82	0,94
Din. O	1,00	1,00	1,00	1,00	0,99	0,96	0,00	0,98	0,97	1,00	1,00	0,99	0,97	0,99
Finlandia	1,00	1,00	1,00	1,00	1,00	0,87	0,98	0,00	0,88	1,00	1,00	0,97	0,48	0,85
Nordpool	1,00	1,00	1,00	1,00	1,00	0,93	0,97	0,88	0,00	1,00	1,00	0,85	0,82	0,88
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,98	0,99	0,97	0,85	1,00	1,00	0,00	0,94	0,95
Suecia	1,00	1,00	1,00	1,00	1,00	0,82	0,97	0,48	0,82	1,00	1,00	0,94	0,00	0,76
Tro	1,00	1,00	1,00	1,00	1,00	0,94	0,99	0,85	0,88	1,00	1,00	0,95	0,76	0,00

Año 2005

<u>V Cramer</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,61	0,69	0,48	0,98	0,98	0,98	0,97	0,98	0,97	0,97	0,97	0,97	0,97
QLD	0,61	0,00	0,82	0,73	0,97	0,98	0,98	0,98	0,97	0,98	0,98	0,97	0,97	0,97
SA	0,69	0,82	0,00	0,62	0,98	0,97	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,97
VIC	0,48	0,73	0,62	0,00	0,98	0,97	0,98	0,98	0,98	0,98	0,98	0,97	0,97	0,98
Austria	0,98	0,97	0,98	0,98	0,00	0,95	0,92	0,96	0,97	0,96	0,98	0,97	0,97	0,97
Din. E	0,98	0,98	0,97	0,97	0,95	0,00	0,90	0,72	0,72	0,97	0,98	0,78	0,58	0,72
Din. O	0,98	0,98	0,98	0,98	0,92	0,90	0,00	0,91	0,90	0,97	0,97	0,91	0,90	0,91
Finlandia	0,97	0,98	0,98	0,98	0,96	0,72	0,91	0,00	0,66	0,97	0,97	0,77	0,44	0,65
Nordpool	0,98	0,97	0,98	0,98	0,97	0,72	0,90	0,66	0,00	0,97	0,98	0,44	0,52	0,56
OMEL	0,97	0,98	0,98	0,98	0,96	0,97	0,97	0,97	0,97	0,00	0,97	0,98	0,98	0,97
Ontario	0,97	0,98	0,98	0,98	0,98	0,98	0,97	0,97	0,98	0,97	0,00	0,99	0,97	0,98
Oslo	0,97	0,97	0,98	0,97	0,97	0,78	0,91	0,77	0,44	0,98	0,99	0,00	0,65	0,66
Suecia	0,97	0,97	0,98	0,97	0,97	0,58	0,90	0,44	0,52	0,98	0,97	0,65	0,00	0,50
Tro	0,97	0,97	0,97	0,98	0,97	0,72	0,91	0,65	0,56	0,97	0,98	0,66	0,50	0,00



Año 2006

<u>V Cramer</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,64	0,65	0,48	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,97
QLD	0,64	0,00	0,83	0,76	0,98	0,97	0,97	0,97	0,98	0,98	0,97	0,98	0,98	0,98
SA	0,65	0,83	0,00	0,53	0,97	0,98	0,97	0,97	0,97	0,97	0,98	0,98	0,97	0,97
VIC	0,48	0,76	0,53	0,00	0,97	0,98	0,98	0,98	0,98	0,97	0,98	0,97	0,97	0,97
Austria	0,98	0,98	0,97	0,97	0,00	0,97	0,94	0,97	0,97	0,97	0,98	0,97	0,97	0,98
Din. E	0,98	0,97	0,98	0,98	0,97	0,00	0,85	0,67	0,79	0,97	0,98	0,90	0,62	0,80
Din. O	0,98	0,97	0,97	0,98	0,94	0,85	0,00	0,87	0,85	0,97	0,97	0,94	0,86	0,92
Finlandia	0,98	0,97	0,97	0,98	0,97	0,67	0,87	0,00	0,70	0,98	0,98	0,84	0,31	0,63
Nordpool	0,98	0,98	0,97	0,98	0,97	0,79	0,85	0,70	0,00	0,98	0,98	0,66	0,64	0,70
OMEL	0,98	0,98	0,97	0,97	0,97	0,97	0,97	0,98	0,98	0,00	0,98	0,98	0,98	0,97
Ontario	0,98	0,97	0,98	0,98	0,98	0,98	0,97	0,98	0,98	0,98	0,00	0,97	0,98	0,98
Oslo	0,98	0,98	0,98	0,97	0,97	0,90	0,94	0,84	0,66	0,98	0,97	0,00	0,77	0,78
Suecia	0,98	0,98	0,97	0,97	0,97	0,62	0,86	0,31	0,64	0,98	0,98	0,77	0,00	0,47
Tro	0,97	0,98	0,97	0,97	0,98	0,80	0,92	0,63	0,70	0,97	0,98	0,78	0,47	0,00

Año 2007

<u>V Cramer</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,69	0,83	0,71	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,69	0,00	0,74	0,66	0,98	0,97	0,98	0,98	0,98	0,97	0,98	0,98	0,98	0,97
SA	0,83	0,74	0,00	0,44	0,98	0,98	0,97	0,98	0,98	0,98	0,98	0,97	0,97	0,98
VIC	0,71	0,66	0,44	0,00	0,98	0,97	0,98	0,98	0,98	0,98	0,98	0,97	0,97	0,98
Austria	1,00	0,98	0,98	0,98	0,00	0,94	0,93	0,96	0,95	0,97	0,97	0,97	0,96	0,97
Din. E	1,00	0,97	0,98	0,97	0,94	0,00	0,77	0,72	0,80	0,97	0,98	0,91	0,69	0,79
Din. O	1,00	0,98	0,97	0,98	0,93	0,77	0,00	0,90	0,85	0,97	0,98	0,96	0,88	0,93
Finlandia	1,00	0,98	0,98	0,98	0,96	0,72	0,90	0,00	0,68	0,97	0,98	0,83	0,19	0,52
Nordpool	1,00	0,98	0,98	0,98	0,95	0,80	0,85	0,68	0,00	0,97	0,97	0,70	0,65	0,69
OMEL	1,00	0,97	0,98	0,98	0,97	0,97	0,97	0,97	0,97	0,00	0,97	0,97	0,98	0,98
Ontario	1,00	0,98	0,98	0,98	0,97	0,98	0,98	0,98	0,97	0,97	0,00	0,98	0,98	0,98
Oslo	1,00	0,98	0,97	0,97	0,97	0,91	0,96	0,83	0,70	0,97	0,98	0,00	0,81	0,80
Suecia	1,00	0,98	0,97	0,97	0,96	0,69	0,88	0,19	0,65	0,98	0,98	0,81	0,00	0,44
Tro	1,00	0,97	0,98	0,98	0,97	0,79	0,93	0,52	0,69	0,98	0,98	0,80	0,44	0,00



Año 2008

V Cramer	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,76	0,86	0,77	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,76	0,00	0,77	0,70	0,98	0,98	0,97	0,97	0,97	0,98	0,98	0,97	0,97	0,98
SA	0,86	0,77	0,00	0,48	0,97	0,97	0,97	0,98	0,97	0,97	0,98	0,98	0,97	0,97
VIC	0,77	0,70	0,48	0,00	0,98	0,97	0,98	0,98	0,98	0,98	0,97	0,98	0,98	0,97
Austria	1,00	0,98	0,97	0,98	0,00	0,92	0,92	0,96	0,95	0,98	0,98	0,97	0,96	0,96
Din. E	1,00	0,98	0,97	0,97	0,92	0,00	0,77	0,73	0,84	0,98	0,98	0,95	0,70	0,86
Din. O	1,00	0,97	0,97	0,98	0,92	0,77	0,00	0,86	0,87	0,97	0,98	0,96	0,85	0,92
Finlandia	1,00	0,97	0,98	0,98	0,96	0,73	0,86	0,00	0,73	0,98	0,98	0,94	0,15	0,67
Nordpool	1,00	0,97	0,97	0,98	0,95	0,84	0,87	0,73	0,00	0,98	0,97	0,82	0,71	0,78
OMEL	1,00	0,98	0,97	0,98	0,98	0,98	0,97	0,98	0,98	0,00	0,98	0,97	0,98	0,98
Ontario	1,00	0,98	0,98	0,97	0,98	0,98	0,98	0,98	0,97	0,98	0,00	0,98	0,98	0,97
Oslo	1,00	0,97	0,98	0,98	0,97	0,95	0,96	0,94	0,82	0,97	0,98	0,00	0,93	0,95
Suecia	1,00	0,97	0,97	0,98	0,96	0,70	0,85	0,15	0,71	0,98	0,98	0,93	0,00	0,61
Tro	1,00	0,98	0,97	0,97	0,96	0,86	0,92	0,67	0,78	0,98	0,97	0,95	0,61	0,00

Año 2009

V Cramer	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,49	0,71	0,61	0,98	0,97	0,97	0,98	0,98	0,98	0,98	0,98	0,98	0,97
QLD	0,49	0,00	0,80	0,71	0,98	0,98	0,97	0,98	0,97	0,97	0,98	0,98	0,98	0,98
SA	0,71	0,80	0,00	0,55	0,98	0,98	0,98	0,97	0,98	0,97	0,97	0,98	0,98	0,98
VIC	0,61	0,71	0,55	0,00	0,97	0,98	0,98	0,97	0,98	0,96	0,97	0,98	0,98	0,97
Austria	0,98	0,98	0,98	0,97	0,00	0,91	0,93	0,93	0,92	0,97	0,97	0,94	0,93	0,95
Din. E	0,97	0,98	0,98	0,98	0,91	0,00	0,80	0,63	0,73	0,97	0,98	0,83	0,58	0,78
Din. O	0,97	0,97	0,98	0,98	0,93	0,80	0,00	0,82	0,77	0,98	0,98	0,87	0,80	0,89
Finlandia	0,98	0,98	0,97	0,97	0,93	0,63	0,82	0,00	0,64	0,97	0,98	0,82	0,18	0,60
Nordpool	0,98	0,97	0,98	0,98	0,92	0,73	0,77	0,64	0,00	0,97	0,97	0,65	0,61	0,69
OMEL	0,98	0,97	0,97	0,96	0,97	0,97	0,98	0,97	0,97	0,00	0,97	0,97	0,97	0,98
Ontario	0,98	0,98	0,97	0,97	0,97	0,98	0,98	0,98	0,97	0,97	0,00	0,97	0,98	0,97
Oslo	0,98	0,98	0,98	0,98	0,94	0,83	0,87	0,82	0,65	0,97	0,97	0,00	0,80	0,82
Suecia	0,98	0,98	0,98	0,98	0,93	0,58	0,80	0,18	0,61	0,97	0,98	0,80	0,00	0,52
Tro	0,97	0,98	0,98	0,97	0,95	0,78	0,89	0,60	0,69	0,98	0,97	0,82	0,52	0,00

A continuación, usando cada matriz de distancias, mostramos el árbol de expansión mínima (mediante el algoritmo de Prim) para los mercados eléctricos en estudio con cada una de las distancias.

Para la distancia V-Cramer se ha calculado el límite de significación para la dependencia y es de 0'97. Es decir, una distancia V-Cramer menor de 0'97 indicará dependencia estadística entre las series y mayor o igual a 0'97 indicará independencia, referido al 95% de confianza.



Distancia V de Cramer

Vértice Origen	Vértice Destino	Distancia
1	4	0,6456
1	2	0,6458
4	3	0,6592
1	9	0,9705
9	13	0,746
13	8	0,355
13	14	0,6334
13	6	0,6803
9	12	0,7575
6	7	0,8859
7	5	0,9519
1	10	0,9707
5	11	0,9773

Distancia Coeficiente de incertidumbre

Vértice Origen	Vértice Destino	Distancia
1	4	0,7479
1	2	0,7501
4	3	0,7594
1	9	0,9957
9	13	0,8226
13	8	0,4783
13	14	0,7602
13	6	0,8151
9	12	0,8499
6	7	0,9586
7	5	0,9916
1	10	0,9959
5	11	0,9981

Distancia Universal 1

Vértice Origen	Vértice Destino	Distancia
1	4	0,7502
1	2	0,7522
4	3	0,76
1	9	0,9964
9	13	0,8233
13	8	0,4807
13	14	0,7613
13	6	0,817
9	12	0,8501
6	7	0,9591
7	5	0,9916
1	10	0,9966
5	11	0,9981

Distancia Universal 2

Vértice Origen	Vértice Destino	Distancia
1	4	0,8558
1	2	0,8572
4	3	0,8633
1	9	0,9979
9	13	0,9027
13	8	0,6471
13	14	0,8638
13	6	0,8982
9	12	0,9189
6	7	0,9788
7	5	0,9958
1	10	0,9979
5	11	0,999

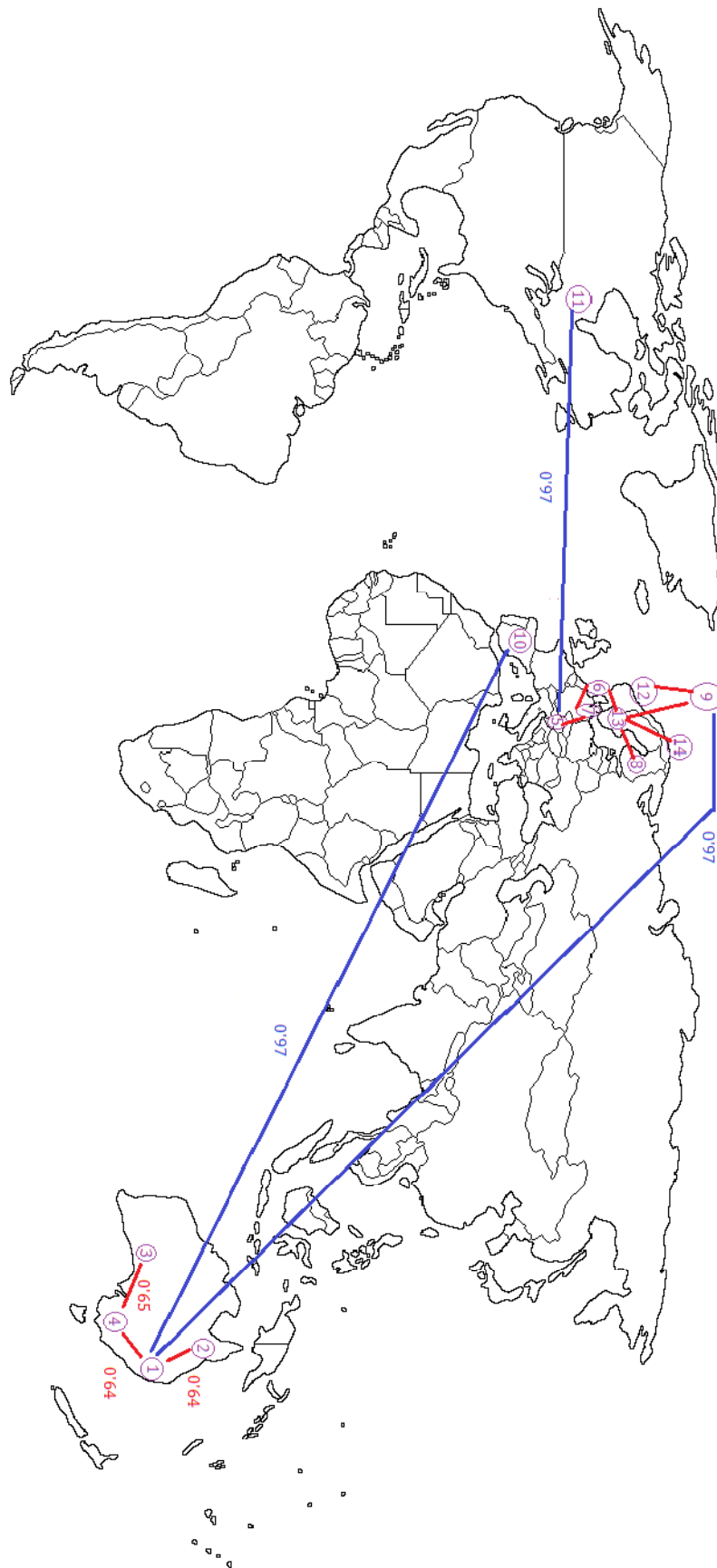


De forma que cada país se corresponde con el número:

1. NSW
2. QLD
3. SA
4. VIC
5. AUSTRIA
6. DIN ESTE
7. DIN OESTE
8. FINLANDIA
9. NORDPOOL
10. OMEL
11. ONTARIO
12. OSLO
13. SUECIA
14. TRO

Obsérvese que el árbol de expansión mínima resultante es el mismo independientemente de la distancia utilizada variando sólo el valor numérico de dichas distancias. En el siguiente gráfico, mostramos las uniones de árbol de expansión mínima colocando cada mercado eléctrico en sus coordenadas geográficas, lo que nos permite apreciar la existencia de clusters homogéneos.

El estudio también se realizó por cada año por separado, obteniéndose resultados similares.



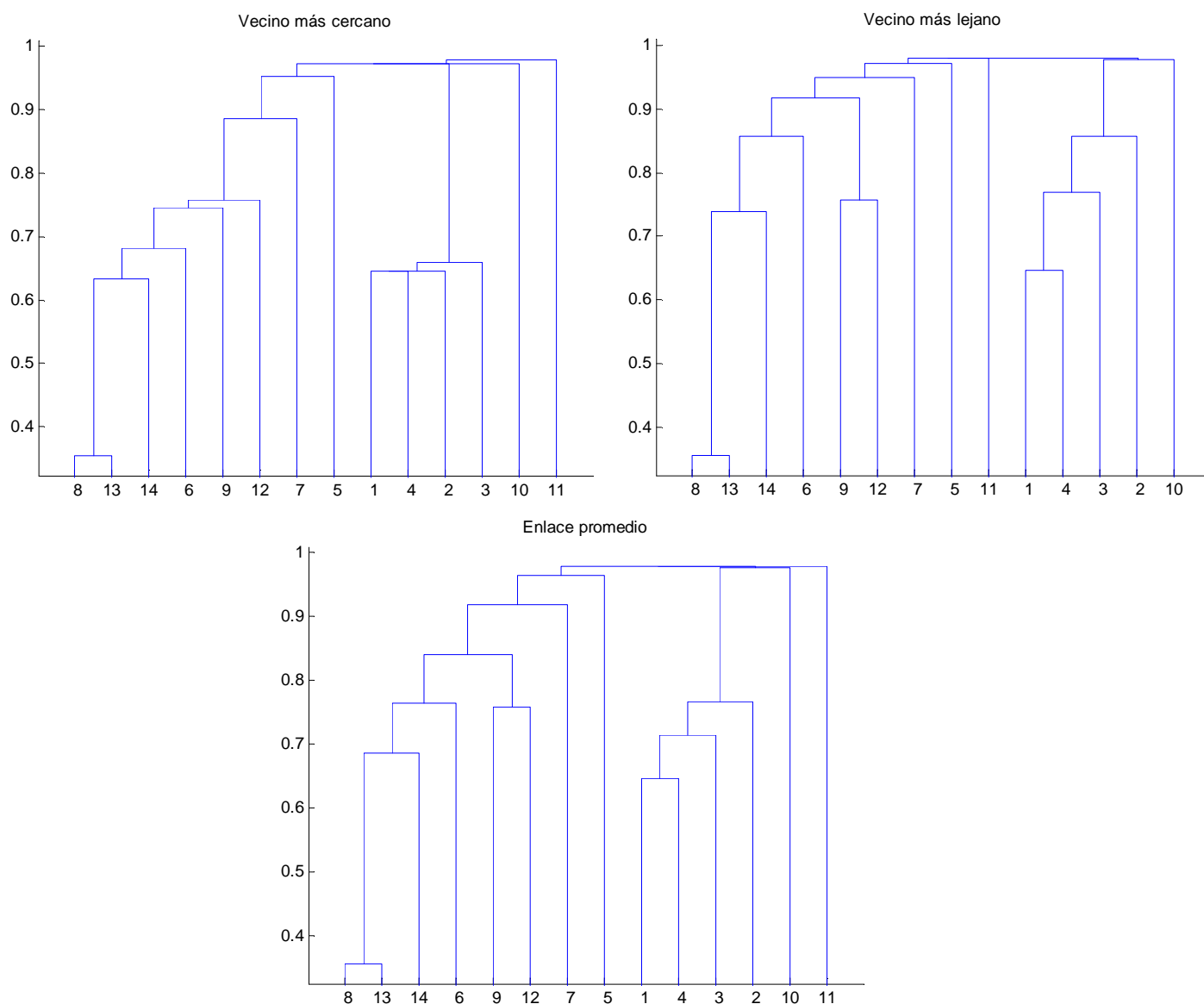
NOTA: En rojo hemos indicado las distancias V-Cramer menores de 0,97 (dependencia estadística) y en azul las distancias mayores o iguales a 0,97 (independencia estadística)



Para finalizar, vamos a realizar una clasificación de los mercados en estudio, obteniendo para ello el dendograma jerárquico. El estudio se ha realizado para cada distancia usando distintos tipos de enlace (vecino más cercano, vecino más lejano y enlace medio entre grupos). Además, se ha tenido en cuenta primero el rango de datos global 2004-2009 y posteriormente año a año.

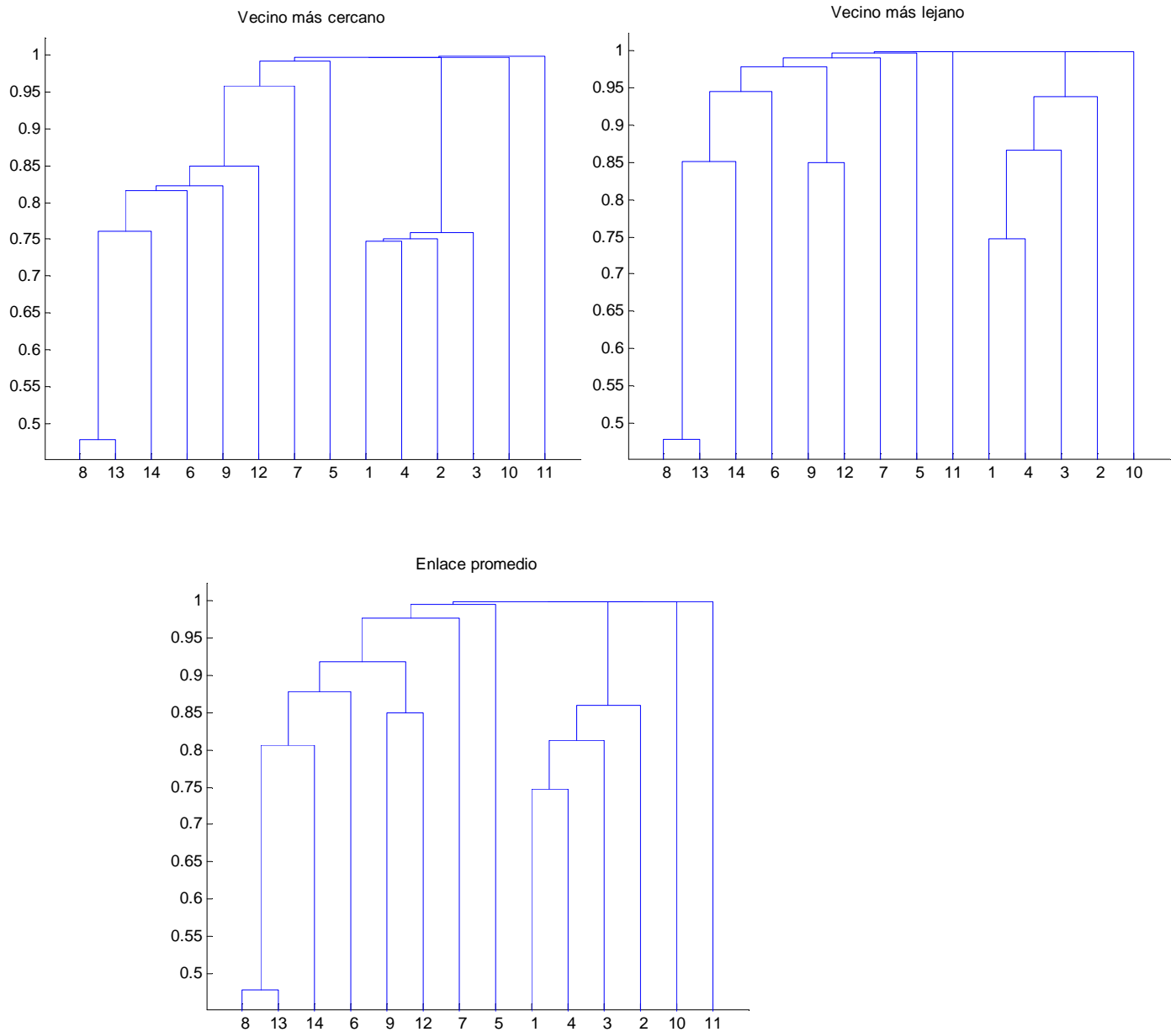
Análisis para rango 2004-2009

- Distancia V de Cramer



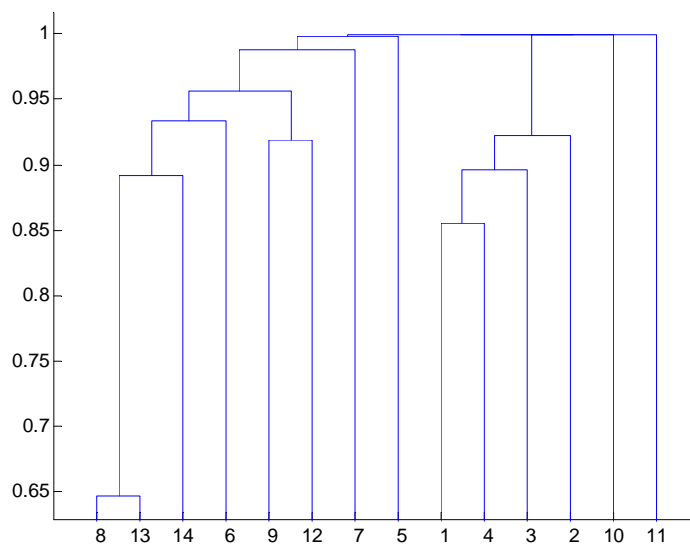
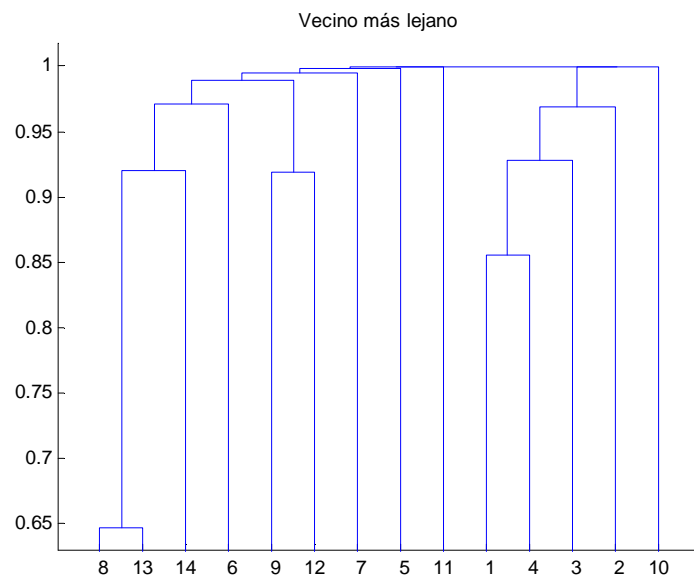
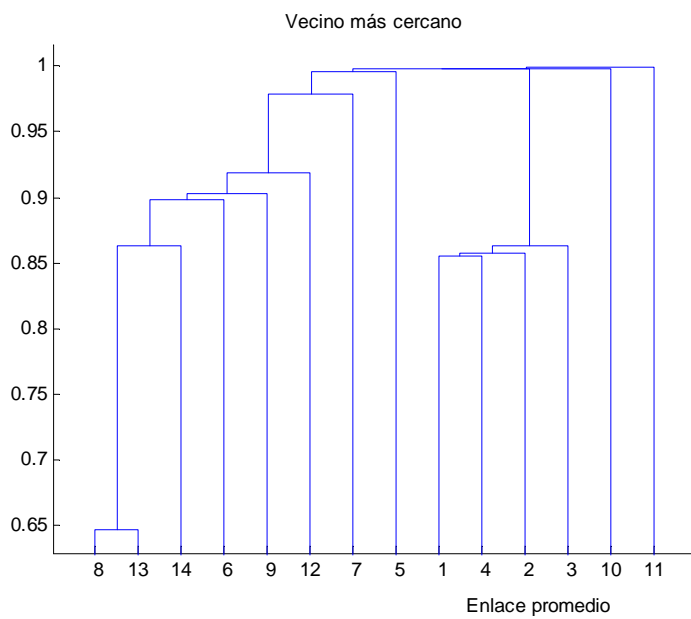


- Distancia Coeficiente Incertidumbre



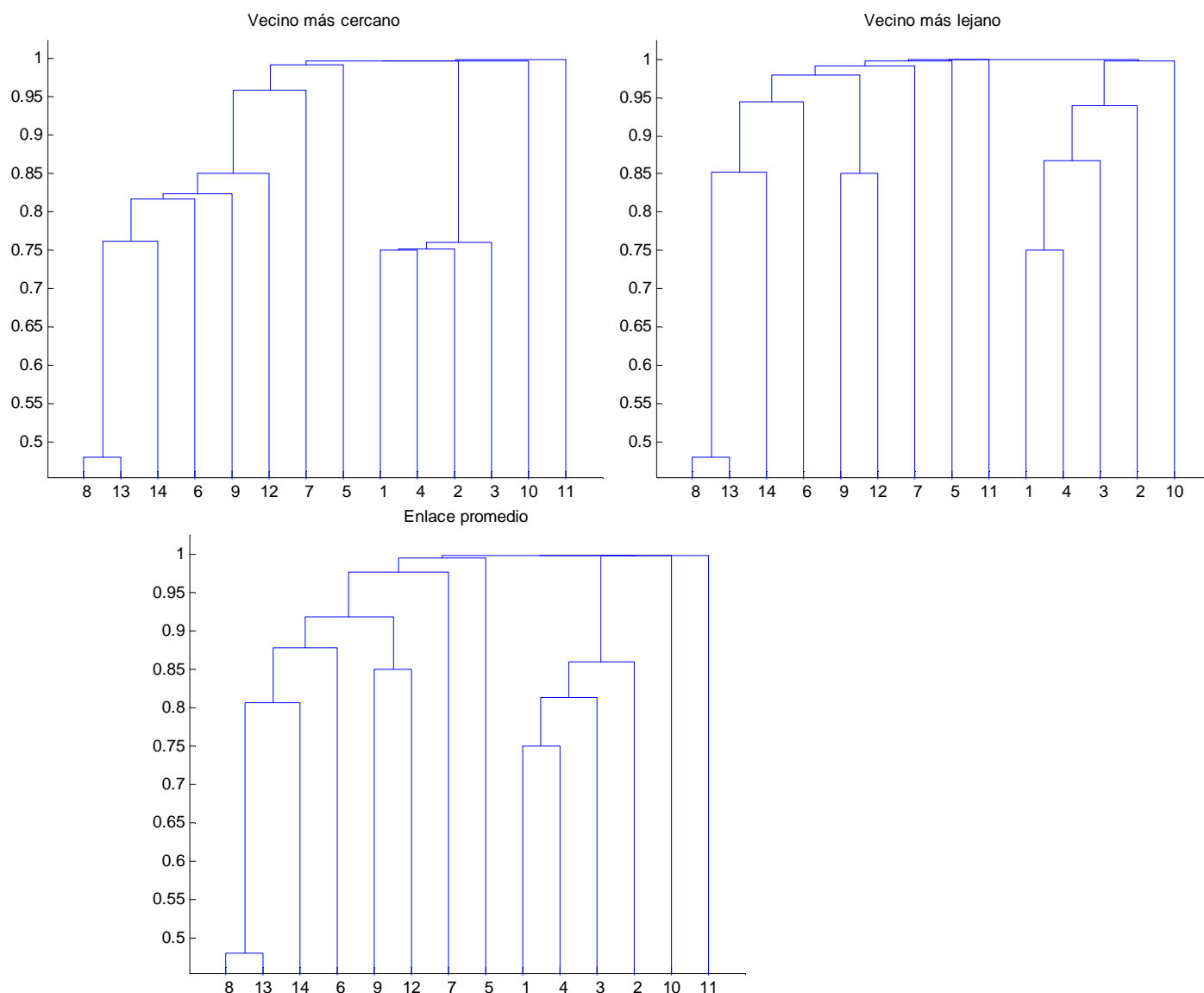


- Distancia Universal 1





- Distancia Universal 2

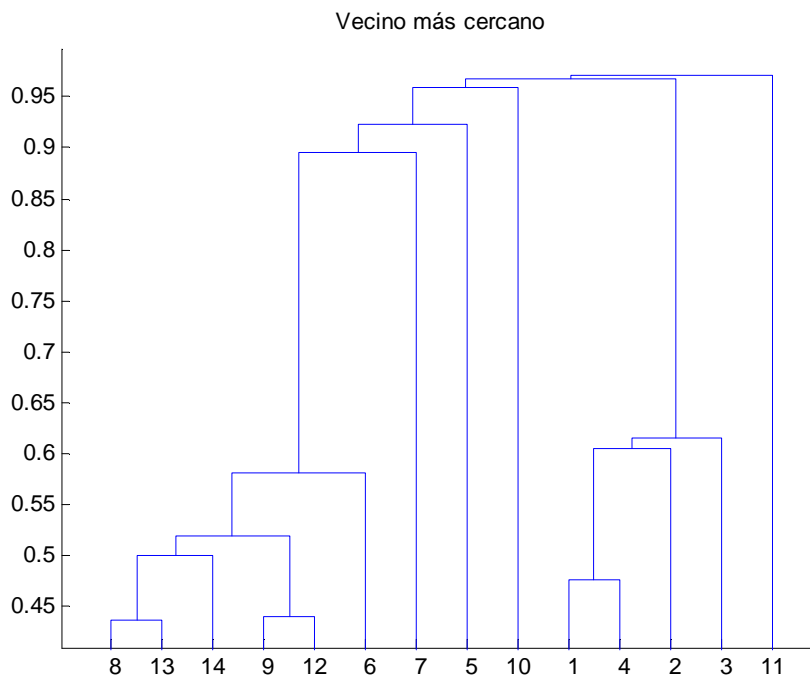


Se observa que los resultados de los dendogramas son casi idénticos independientemente del tipo de enlace y de la distancia usada. Para el análisis año a año no se mostrarán todos los dendogramas, sino aquellos que muestren algún dato relevante.

Atendiendo a los dendogramas, se pueden apreciar de forma clara 4 clusters:

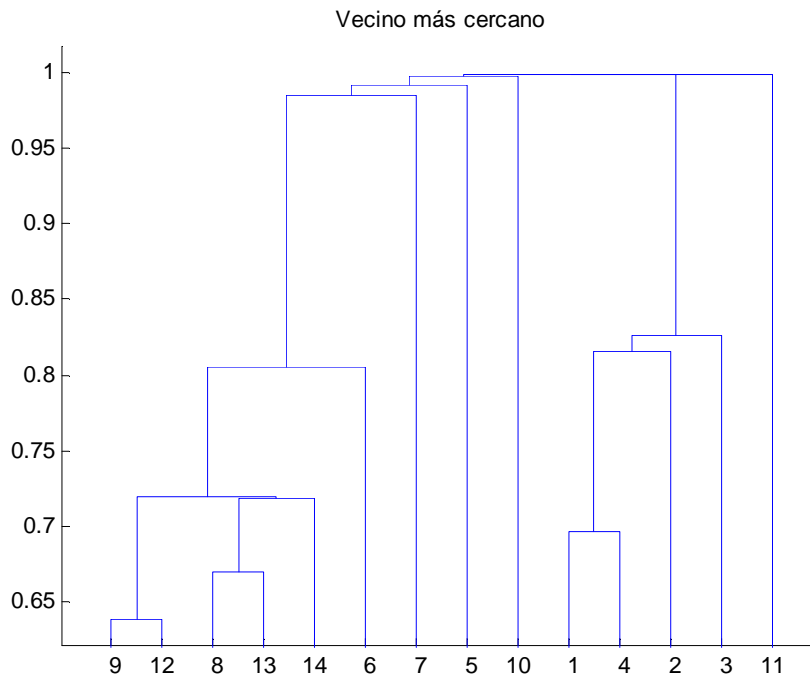
- Clúster 1: Formado por las 4 zonas de Australia (Victoria, Nueva Gales del Sur, Australia del Sur y Queensland).
- Clúster 2: Formado por las distintas regiones del Nordpool (sistema, Finlandia, Suecia, Tro, Oslo y Dinamarca Este) junto con Austria. Resulta interesante ver que Dinamarca Oeste no tiene una gran relación con el resto de países nórdicos.
- Clúster 3: formado únicamente por OMEL.
- Clúster 4: formado únicamente por Ontario.

Análisis para el año 2005



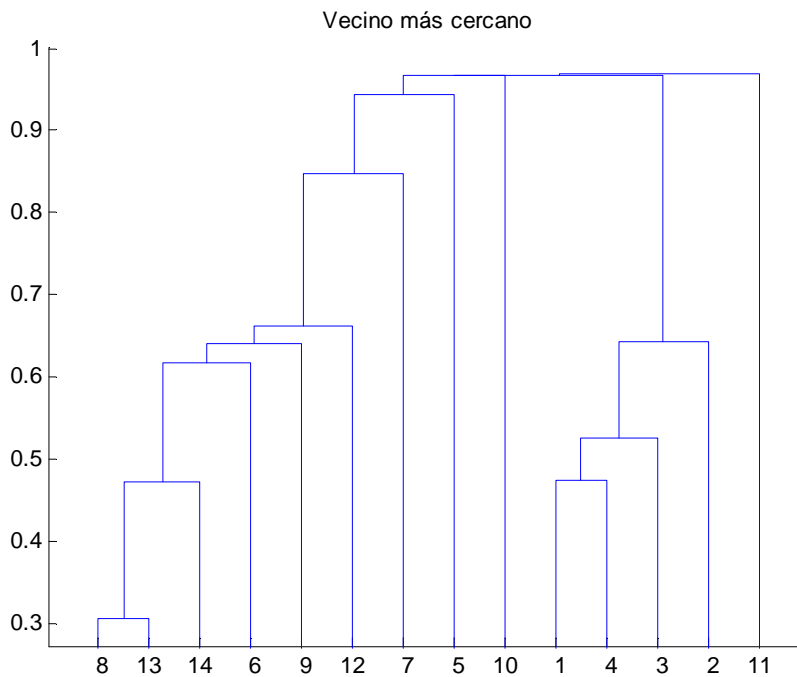
El dendrograma muestra las distancia V de Cramer utilizando el enlace de vecino más cercano. En este caso, podríamos distinguir 3 clusters diferenciados:

- Clúster 1: Formado por las cuatro zonas de Australia, donde destaca la relación entre Nueva Gales del Sur y Australia del Sur.
- Clúster 2: Formados por las distintas regiones de Nordpool. En este caso, la relación existente entre Finlandia y Suecia es tan estrecha como la que hay entre sistema y Oslo. También dentro de este clúster, pero aumentando la heterogeneidad, estarían Dinamarca Oeste, Austria y OMEL. A diferencia del análisis global de 2004-2009, el mercado español en 2005 presentó algo más de relación con sus vecinos europeos.
- Clúster 3: Formado únicamente por Ontario, lo que indicaría que la relación de este mercado con el resto es nula.



El anterior gráfico corresponde a la distancia del Coeficiente de Incertidumbre con el vecino más cercano. La única diferencia respecto a la distancia V de Cramer, es que en este caso los mercados que tienen una relación más fuerte son sistema y Oslo. El resto de distancias y de enlaces proporcionan resultados muy similares (véase Anexo).

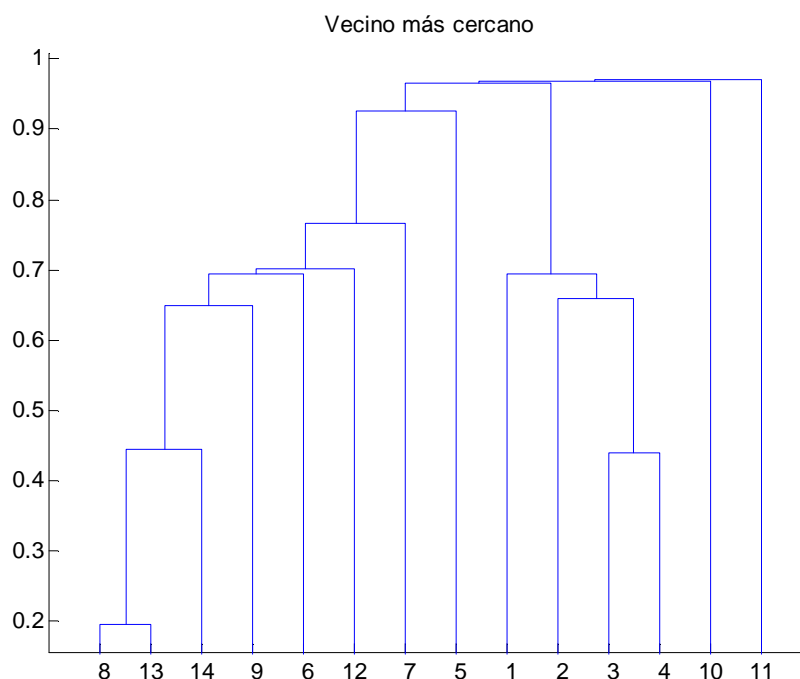
Análisis para año 2006





Este gráfico nos muestra la distancia V-Cramer usando el enlace más cercano. Los resultados para el año 2006 presentan una diferencia principal respecto al año anterior: la leve asociación del mercado español con los mercados europeos en 2005 se hace nula para el año 2006. Además la relación entre los precios del sistema Nordpool y Oslo se hace más débil que el año anterior.

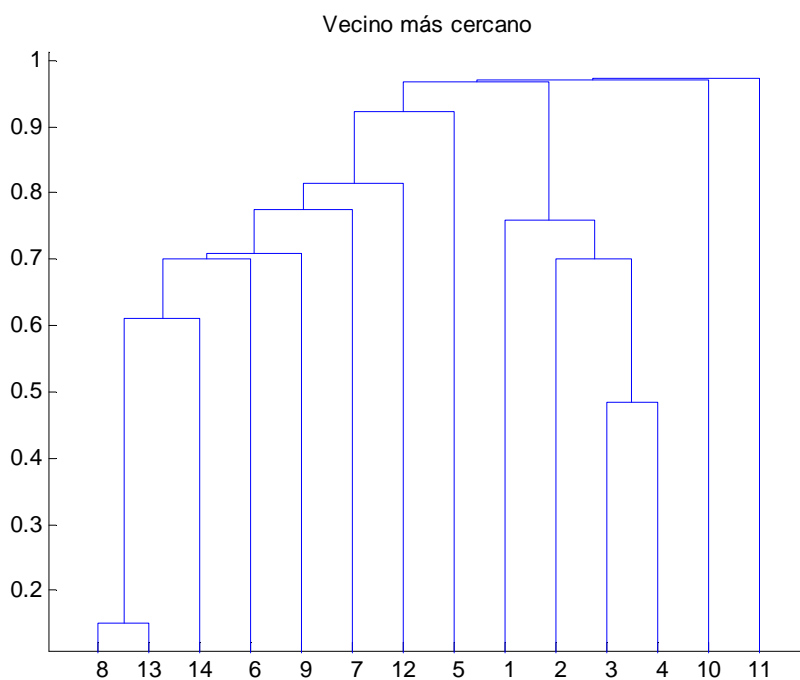
Análisis para año 2007



En 2007 se ha representando la distancia V-Cramer y se observa que en el mercado australiano, las zonas que presentan más dependencia son Australia del Sur y Victoria. Destacar, que este año Dinamarca Oeste tiene una relación más estrecha con los demás países nórdicos que los años anteriores. La clasificación que muestra el dendograma coincide con la obtenida para el rango de años 2004-2009, donde OMEL y Ontario forman clusters individuales.



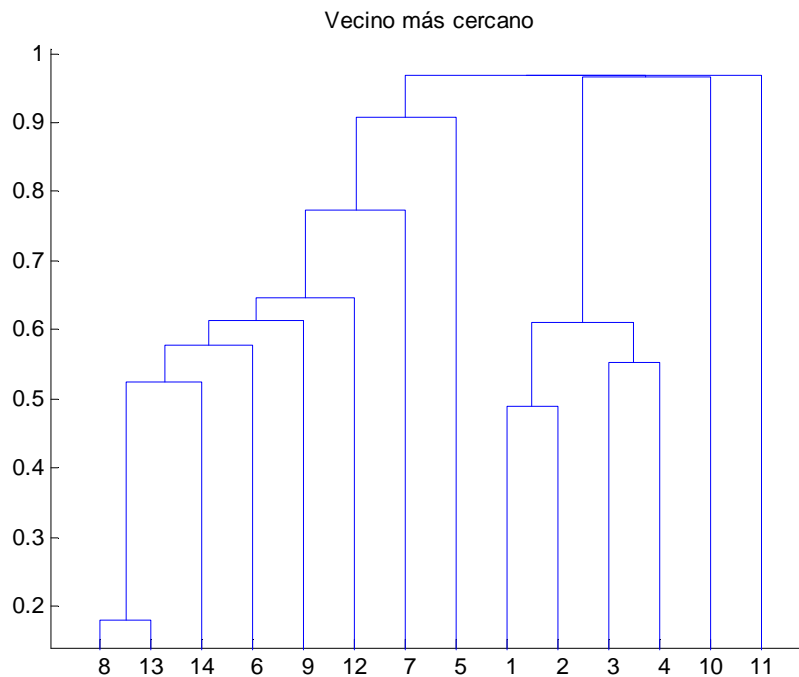
Análisis para año 2008



Una vez más se ha representado la distancia V-Cramer para el año 2008, en el que se repite la situación: OMEL y Ontario aparecen como clusters individuales, Austria aparece incluida en el clúster de los países nórdicos y las 4 zonas de Australia se unen en el clúster más homogéneo.



Análisis para año 2009



Los resultados mostrados, usando la distancia V-Cramer, para el año 2009 son similares a los de 2007, es decir, se aprecian 4 clusters, donde OMEL y Ontario forman clusters individuales.



6. CONCLUSIONES Y LÍNEAS FUTURAS

En este Proyecto, proponemos una metodología para estudiar la asociación entre pares de series temporales, no sólo de manera global, sino por ventanas móviles que se pueden corresponder con un periodo temporal de interés (meses, estaciones, años, etc.)

Mediante datos simulados, hemos analizado la eficiencia del método, mostrando las ventajas frente a otras medidas usuales como es la correlación de Pearson o de Spearman. Para la detección de dependencias, se han codificado las series usando permutaciones y se ha considerado la tabla de contingencia resultante, con los estadísticos usuales Ji-cuadrado de Pearson y razón de verosimilitudes (o G-test).

Sin embargo, estos estadísticos presentan algunas limitaciones, que hemos intentado solventar mediante una adecuada selección de los parámetros embedding (m) y tamaño de ventana móvil (w).

Por otra parte, se ha destacado que el estadístico Ji-cuadrado es sensible al tamaño muestral a la hora de contrastar la independencia. Con el fin de minimizar el efecto del tamaño muestral, así como de tener medidas de asociación y de “distancia” acotadas entre cero y uno, se han propuesto los siguientes coeficientes: medida de asociación V de Cramer, distancia V de Cramer, medida de asociación Coeficiente de Incertidumbre, distancia Coeficiente de Incertidumbre, distancia Universal(1) y distancia Universal(2). Estas medidas propuestas permiten cuantificar el grado de asociación o discrepancia (distancia) entre series, independientemente del tamaño muestral, dimensión embedding y tamaño de ventana móvil.

Se ha utilizado los estadísticos Ji-cuadrado, G-test y Cressie-Read cuando el objetivo es contrastar la hipótesis de independencia entre series, puesto que disponemos de la distribución asintótica que permite establecer las regiones de aceptación y rechazo. Sin embargo, cuando el objetivo sea medir la fuerza de la asociación y que sea comparable entre pares de series, usaremos las medidas de asociación y distancias normalizadas.

Por una parte, destacar la gran efectividad que presentan los estadísticos Ji-cuadrado, G-test y Cressie-Read en la detección de dependencias, que ha sido testada mediante datos simulados. Por otra parte, hemos destacado la detección de “falsas” dependencias debidas a la presencia de componente estacional, lo pone en evidencia la necesidad de eliminar dicha componente antes de estudiar la asociación entre las series. Cuando se trate de datos horarios, como es el caso de los precios de los mercados eléctricos, debemos tener en cuenta dos estacionalidades: semanal y diaria. Por último, indicar un “inconveniente” del método a la hora de analizar autodependencias: cualquier serie (incluyendo un ruido o perturbación aleatoria), va a presentar dependencia con su retardo de orden 1, de orden 2, ...hasta orden ($m-1$), siendo m la dimensión embedding. Esta situación ocurre por la naturaleza del método propuesto.



La aplicación de la metodología propuesta a mercados eléctricos nos ha permitido identificar: fuertes asociaciones entre las series de precios de algunos mercados, asociaciones débiles y ausencia de asociación. Conviene destacar que, en este sentido, los precios del mercado español (OMEL) parecen comportarse como una isla geográfica, sin relación significativa con el resto de mercados europeos analizados. Además, la obtención del árbol de expansión mínima y de los dendogramas para distintas distancias, permiten realizar una clasificación de los mercados analizados en grupos homogéneos.

Cabe destacar la utilidad del software desarrollado para la aplicación de la metodología propuesta en otros contextos, que permitirá calcular los estadísticos mencionados para series de datos de gran longitud (tanto los datos simulados como los precios de los distintos mercados son series de unos 100.000 datos). Se proporciona el código para aquella persona que desee seguir incrementando las funcionalidades de los distintos programas.

Entre las líneas futuras podemos indicar las siguientes.

- Completar el estudio propuesto en este Proyecto incluyendo datos de nuevos mercados eléctricos, a ser posible de distintos continentes.
- Analizar la dependencia entre los precios de la electricidad y otros factores, como pueden ser los consumos globales del mercado, la climatología, las reservas de agua de la zona (para los mercados con un alto porcentaje de generación de hidráulica).
- Estudiar la dependencia entre los precios de la electricidad y los precios de otras fuentes energéticas: gas natural, petróleo, etc.
- Analizar la dependencia entre los precios de la electricidad y otros factores socioeconómicos: cambio dólar/euro, cambios en la regulación del mercado, etc.

La realización de este tipo de análisis permitirá identificar de forma adecuada los factores que influyen en la determinación del precio de la electricidad, consiguiendo una mejora en los modelos de previsión de precios. mercados, sino analizar la relación del precio con otras energías.



7. BIBLIOGRAFÍA

- [1] Christoph Bandt and Bernd Pompe. “Permutation entropy- a natural complexity measure for time series”. Institute of Mathematics and Institute of Physics, University of Greifswald, Germany.
- [2] Yinhe Cao, Wen-wen Tung, J. B. Gao, V. A. Protopopescu y L. M. Hively. “Detecting dynamical changes in time series using the permutation entropy”.
- [3] J. S. Cánovas, A. Guillamón y M. C. Ruiz. “Using permutations to find structural changes in time series”. Universidad Politécnica de Cartagena.
- [4] Matesanz, David y Ortega, Guillermo J. “Network analysis of exchange data: Interdependence drives crisis contagion”. Universidad de Oviedo, Universidad Nacional de Quilmes y CONICET.
- [5] Miryam Salvador. “Análisis del mercado eléctrico en España”. Universidad Politécnica de Cataluña.
- [6] Vincent Grégoire, Christian Genest y Michel Gendron. “Using copulas to model price dependence in energy markets”.
- [7] Iva Mihaylova. “Stochastic dependencies of spot prices in the European electricity market”. Universidad de St. Gallen.
- [8] David Erni. “Cointegration in Spot Price Energy Markets, an analysis using the Kalman filter”. Universidad de St. Gallen.
- [9] Enric Gironés Sola. “Métodes de previsió de volatilitat a curt termini aplicats al Mercat Eléctric Espanyol”. Universidad Politécnica de Cataluña.
- [10] Hamidreza Zareipour. “Price Forecasting and Optimal Operation of Wholesale Customers in a Competitive Electricity Market”. Universidad de Waterloo, Ontario.
- [11] Ángel León y Antonio Rubia. “Comportamiento del precio y volatilidad en el pool eléctrico español”. Universidad de Alicante.
- [12] Alexander Kraskov, Harald Stögbauer, Ralph G. Andrzejak, y Peter Grassberger. “Hierarchical Clustering Based on Mutual Inform”. John-von-Neumann Institute for Computing.
- [13] Unai Aldasoro Marcellan. “Optimización de modelos estocásticos de mercado eléctrico múltiple mediante métodos duales”. Universidad Politécnica de Cataluña.
- [14] María Eugenia Ruiz Molina. “Liberalización del mercado eléctrico y elegibilidad: consecuencias para el consumidor”.
- [15] F. J. Barón López y F. Téllez Montiel. “Capítulo 7: Independencia de variables categóricas”. Apuntes de Bioestadística. Universidad de Málaga.
- [16] José Vicéns Otero y Eva Medina Moral. “Análisis de datos cuantitativos”.
- [17] Sean Wallis. “Measures of association for contingency tables”. University College London.
- [18] Manual StatXact. “Cap. 17: Introduction to Exact Nonparametric Inference”.
- [19] Ana María Aguilera del Pino. “Análisis de tablas de contingencia bidimensionales”. Universidad de Granada.



- [20] Elkin Castaño, Jorge Martínez. “Uso de la función de correlación cruzada en la identificación de modelos ARIMA”. Universidad Nacional de Medellín, Colombia.
- [21] Rafael Weron. “Modeling and Forecasting Electricity Load and Prices: A Statistical Approach”.
- [22] José Miguel Rodríguez Gómez. “Entropías topológica y de permutación. Aplicación a la dinámica económica”. Universidad Politécnica de Cartagena.
- [23] Ley del Sector Eléctrico 54- 1997 27 de noviembre.
- [24] Javier García de Jalón, José Ignacio Rodríguez y Jesús Vidal. “Aprenda Matlab 7.0 como si estuviera en primero”. Universidad Politécnica de Madrid.
- [25] José Carlos Fernández Pérez. “Análisis y evaluación de mercados eléctricos liberalizados a escala internacional”. Universidad Pontificia de Comillas.
- [26] www.ree.es/operacion
- [27] Yinhe Cao, Wen-wen Tung, J. B. Gao, V. A. Protopopescu y L. M. Hively. “Detecting dynamical changes in time series using the permutation entropy”.



I. ANEXO

a. Programación Matlab

i. Función apilar

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [todo]=apilar(frec,m)
n=factorial(m);
for j=1:n
for t=1:n
col((j-1)*n+t)=t;
fil((t-1)*n+j)=t;
datos2((j-1)*n+t)=frec(t,j);
end
end
col=col';
fil=fil';
datos2=datos2';
todo=[datos2 col fil];
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

ii. Función desestacionalizar

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [tabladesest]=desestacionalizar(serie)

x = serie;
px = x(169:size(x));
qx = x(1:size(x)-168);
y = px-qx;

py = y(25:size(y));
qy = y(1:size(y)-24);
z = py-qy;

suma_sem=zeros(length(x)-840,1);
suma_sem=suma_sem';
suma_dias=suma_sem;
suma_ds=suma_sem;

mx(1,:)=x(841:size(x));
for i=1:1:4
mx(i+1,:)=x(840-168*i+1:size(x)-168*i);
suma_sem = suma_sem+mx(i+1,:);
end
for i=1:1:7
mx(i+5,:)=x(840-24*i+1:size(x)-24*i);
suma_dias = suma_dias+mx(i+5,:);
end
for i=8:1:35
mx(i+5,:)=x(840-24*i+1:size(x)-24*i);
suma_ds = suma_ds+mx(i+5,:);
end
zweron = mx(1,:)-((1/4)*suma_sem+(1/7)*suma_dias-(1/28)*suma_ds);
```



```
menor=min([length(x) length(y) length(z) length(zweron)]);

tabladesest(1,:)=x((length(x)-menor+1):length(x));

tabladesest(2,:)=y((length(y)-menor+1):length(y));

tabladesest(3,:)=z((length(z)-menor+1):length(z));

tabladesest(4,:)=zweron((length(zweron)-menor+1):length(zweron));

end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

iii. Función ordenar

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function permuta = ordenar(serie,m)

tam=length(serie);
longitud=tam-m+1;

for t=1:longitud

    serieaux = serie(t:t+m-1);

    [a,b] = sort(serieaux);           %ordena serieaux

    permuta(t,:)=b;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

iv. Función prim

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [mst, cost]=prim(A)

[n,m]=size(A); % Matriz de tamaño n por n, donde n = # nodos.
display(A);
%
if norm(A-A','fro')~= 0 , % Si la matriz de adyacencia no es simetrica
    disp(' Error: La matriz de Adyacencia debe ser simétrica ')
    return,
end

intree=1;
number_in_tree=1;
number_of_edges = 0;
notinintree=(2:n)';
number_notin_tree=n-1;
in=intree(1:number_in_tree); % imprime los nodos que están en el árbol
display(in);
out=notinintree(1:number_notin_tree);
display(out)
%pause, % y cuales no.
```




```
% Iterar hasta que todos los n nodos están en el árbol.

while number_in_tree<n

    mincost = Inf; % You can actually enter infinity into Matlab.
    for i=1:number_in_tree
        for j=1:number_notin_tree,
            ii = intree(i);
            jj = notintree(j);
            if A(ii,jj)< mincost && A(ii,jj)~=0
                mincost=A(ii,jj);
                jsave=j;
                iisave=ii;
                jjsave=jj; % Save coords of node.
            end
        end
    end

    number_of_edges = number_of_edges+1;
    mst(number_of_edges,1)=iisave;
    mst(number_of_edges,2)=jjsave;
    costs(number_of_edges,1)=mincost;

    number_in_tree=number_in_tree+1;
    intree=[intree;jjsave]; % Add this node to tree.
    for j=jsave+1:number_notin_tree,
        notintree(j-1) = notintree(j);
    end;
    number_notin_tree = number_notin_tree-1;
    in=intree(1:number_in_tree);
    display(in);
    out=notintree(1:number_notin_tree);
    display(out);
    %pause,% which are not.

end
disp(' Aristas en el árbol de expansión mínimo y sus costos ')
display([mst costs]) % Print out edges in minimum spanning tree.
cost=sum(costs);
display(cost);
disp('Aristas que Se Unen')
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

v. Función tablabidim2

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [tabla]=
tablabidim2(serie1,serie2,m,ventana,retardo,landa,fecha1,fecha2)

serie1=serie1(1+fecha1:length(serie1));
factorialm=factorial(m);
%serie2=serie2(1+fecha:length(serie2)); % viejo
serie2=serie2(1+fecha2+retardo:length(serie2));

ni=0;
nj=0;
```



```
if m==3
    limitem=37.65;
elseif m==4
    limitem=583.61;
end

if length(serie1) <= length(serie2)
    a=length(serie1);
else
    a=length(serie2);
end

nchi=fix(a/ventana);
% inicio=0;
% fin=0;
for i=1:nchi

    %     inicio=fin+1;
    %     fin=inicio+ventana-1;
    %     if i==nchi
    %         fin=a;
    %     end
    if i==1
        inicio=1;
        fin=inicio+ventana-1;
    %     elseif i==nchi
    %         inicio=ventana*(i-1)+i-1;
    %         fin=a-retardo;
    %         ultimos datos no se quieren
    %         se pierde porque los

    else

        % inicio=ventana*(i-1)+i-1; viejo
        inicio=ventana*(i-1)+1;
        fin=inicio+ventana-1;

    end

    serie1aux = serie1(inicio:fin);
    %     serie2aux = serie2(inicio+retardo:fin+retardo);
    serie2aux = serie2(inicio:fin);
    %     viejo

    aux=perms(1:m); %posibles permutaciones
    indice=(1:factorialm);
    tab1ax=[aux indice'];
    tab1ay=[aux indice'];
    frec=zeros(factorialm,factorialm); %crea matriz de ceros)
    tam=min(length(serie1aux),length(serie2aux));

    longitud=tam-m+1;
    permutax = ordenar(serie1aux,m);
    permutay = ordenar(serie2aux,m);

    for t=1:longitud
```



```
for t1=1:factorialm
    if permutax(t,:)==tablax(t1,1:m)
        for t2=1:factorialm
            if permutay(t,:)==tablay(t2,1:m)
                frec(t1,t2)=frec(t1,t2)+1;
                break; %puesto lo ultimo
            end
        end
        break; %puesto tb lo ultimo
    end
end
end
todo=apilar(frec,m);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
indice2=0;
indice1=0;
for t1=1:factorialm
    if sum(frec(t1,:))~=0
        indice1=indice1+1;
        almacen1(indice1)=t1;
    end
    if sum(frec(:,t1))~=0
        indice2=indice2+1;
        almacen2(indice2)=t1;
    end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% FRECUENCIAS:
%           frec_esperada -> FRECUENCIA ESPERADA
%           frec_obs      -> FRECUENCIA OBSERVADA
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
total_obs=sum(sum(frec));
for t1=1:indice1
    for t2=1:indice2
        tabla_new(t1,t2)=frec(almacen1(t1),almacen2(t2));
        frec_obs(t1,t2)=frec(almacen1(t1),almacen2(t2));
        tabla(20+t1,t2)=frec_obs(t1,t2);
    end
end

for t1=1:indice1
    for t2=1:indice2

frec_esperada(t1,t2)=sum(tabla_new(t1,:))*sum(tabla_new(:,t2))/total_o
bs;

    end
end
grad_lib=(indice1-1)*(indice2-1);

%Cálculo del estadístico CHI2 sin interface R
suma_chi=0;
cr=0;
gtest=0;
for t1=1:indice1
    for t2=1:indice2
        suma_chi=suma_chi+(frec_obs(t1,t2)-
frec_esperada(t1,t2))^2/frec_esperada(t1,t2);

cr=cr+frec_obs(t1,t2)*((frec_obs(t1,t2)/frec_esperada(t1,t2))^landa-
1);
```



```
        if frec_obs(t1,t2)~=0

gtest=gtest+frec_obs(t1,t2)*log(frec_obs(t1,t2)/frec_esperada(t1,t2));
        end

    end
end

k=factorialm;

% tcramer=fecha2-fecha1;

ncramer=ventana-m+1;

vcramer=sqrt(suma_chi/(ncramer*(k-1)));
dvcramer=1-vcramer;

hxy=0;
for q=1:factorialm
    for t=1:factorialm
        ni=ni+frec_obs(q,t);
        nj=nj+frec_obs(t,q);
        if frec_obs(t,q)~=0 %MODIFICADO MCRUIZ
            hxy=hxy-
(frec_obs(t,q)/ncramer)*log(frec_obs(t,q)/ncramer);
        end
    end

    tabla(14,q)=ni;
    tabla(13,q)=nj;
    ni=0;
    nj=0;
end

hx=0;
hy=0;
for q=1:factorialm
    if tabla(14,q)~=0 %MODIFICADO MCRUIZ
        hx=hx-(tabla(14,q)/ncramer)*log(tabla(14,q)/ncramer);
    end
    if tabla(13,q)~=0 %MODIFICADO MCRUIZ
        hy=hy-(tabla(13,q)/ncramer)*log(tabla(13,q)/ncramer);
    end
end

ixy=hx+hy-hxy;
d1=1-(ixy/hxy);

if hx>=hy
    d2=1-(ixy/hx);
else
    d2=1-(ixy/hy);
end
```



```
% dixy=1-ixy; %MODIFICADO MCRUIZ
dixy=1-(2*ixy/(hx+hy)); %MODIFICADO MCRUIZ

gtestfinal=gtest*2;
[RHO1,PVAL1]=corr(serie1aux,serie2aux,'type','Spearman');
[RHO2,PVAL2]=corr(serie1aux,serie2aux,'type','Pearson');
%cambiar spearman por
% pearson y lo calcula
%elnuevoframe= PVAL;
%[RHO,PVAL]=corr(serie1aux,'type','Spearman');
crfinal=cr*2/(landa*(landa+1));
% val_chi2=suma_chi;
% val_crit_chi=chi2inv(0.95,grad_lib);
% p_valor_chi=1-cdf('chi2',val_chi2,grad_lib);
% display(['Estadístico Chi2 : ',num2str(val_chi2)]);
% display(['p-valor Chi2: ',num2str(p_valor_chi)]);
tabla(1,i)=suma_chi;
tabla(2,i)=crfinal;
tabla(3,i)=gtestfinal;
tabla(4,i)=RHO1;
tabla(5,i)=RHO2;
tabla(6,i)=PVAL1;
tabla(7,i)=PVAL2;
tabla(8,i)=37.65;
tabla(9,i)=583.61;
tabla(10,i)=0.05;
tabla(11,i)=m;
tabla(15,i)=dvcramer;
tabla(16,i)=dixy;
tabla(17,i)=d1;
tabla(18,i)=d2;

if i==1

    for j=1:5
        tabla(12,j)=0;
    end
end

if m==3 || m==4

    for z=1:3

        if tabla(z,i)<limitem
            tabla(12,z)=tabla(12,z)+1;
        end
    end
end

%
%
if tabla(6,i)>0.05
    tabla(12,4)=tabla(12,4)+1;
end

if tabla(7,i)>0.05
    tabla(12,5)=tabla(12,5)+1;
```



```

end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

vi. Función tablabidretardo

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [tabla]=
tablabidretardo(serie1,serie2,m,retardoi,retardof,landa,fechali,fechal
f,fecha2i,fecha2f)

ni=0;
nj=0;

factorialm=factorial(m);

k=factorialm;

if m==3
    limitem=37.65;
elseif m==4
    limitem=583.61;
end

for i=1:(retardof-retardoi+1)

    serie1aux=serie1(fechali:fechalf-i);
    %serie1aux=serie1(1+fechali:fechalf-i+1);
    permutax=ordenar(serie1aux,m);
    serie2aux=serie2(fecha2i+retardoi+i-1:fecha2f+retardoi-1);
    %serie2aux=serie2(fecha2i+retardoi+i:fecha2f);!

    aux=perms(1:m); %posibles permutaciones
    indice=(1:factorialm);
    tablux=[aux indice']; %pa q le pone el indice??
    tablay=[aux indice'];
    frec=zeros(factorialm,factorialm); %crea matriz de ceros)
    tam=min(length(serie1aux),length(serie2aux));

    longitud=tam-m+1;

    permutay = ordenar(serie2aux,m);

    for t=1:longitud
        for t1=1:factorialm
            if permutax(t,:)==tablux(t1,1:m)
                for t2=1:factorialm
                    if permutay(t,:)==tablay(t2,1:m)
                        frec(t1,t2)=frec(t1,t2)+1;
                        break; %puesto lo ultimo
                    end
                end
            end
        end
    end
end

```



```
end
break; %puesto tb lo ultimo
end

end

end
todo=apilar(frec,m);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
indice2=0;
indice1=0;
for t1=1:factorialm
    if sum(frec(t1,:))~=0
        indice1=indice1+1;
        almacen1(indice1)=t1;
    end
    if sum(frec(:,t1))~=0
        indice2=indice2+1;
        almacen2(indice2)=t1;
    end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% FRECUENCIAS:
%         frec_esperada -> FRECUENCIA ESPERADA
%         frec_obs      -> FRECUENCIA OBSERVADA
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
total_obs=sum(sum(frec));
for t1=1:indice1
    for t2=1:indice2
        tabla_new(t1,t2)=frec(almacen1(t1),almacen2(t2));
        frec_obs(t1,t2)=frec(almacen1(t1),almacen2(t2));
    end
end

for t1=1:indice1
    for t2=1:indice2

frec_esperada(t1,t2)=sum(tabla_new(t1,:))*sum(tabla_new(:,t2))/total_o
bs;

end
end
grad_lib=(indice1-1)*(indice2-1);

%Cálculo del estadístico CHI2 sin interface R
suma_chi=0;
cr=0;
gtest=0;
for t1=1:indice1
    for t2=1:indice2
        suma_chi=suma_chi+(frec_obs(t1,t2)-
frec_esperada(t1,t2))^2/frec_esperada(t1,t2);

cr=cr+frec_obs(t1,t2)*((frec_obs(t1,t2)/frec_esperada(t1,t2))^landa-
1);
        if frec_obs(t1,t2)~=0

gtest=gtest+frec_obs(t1,t2)*log(frec_obs(t1,t2)/frec_esperada(t1,t2));
end

end
end
```



```
ncramer=length(serie1aux)-m+1;

vcramer=sqrt(suma_chi/(ncramer*(k-1)));
dvcramer=1-vcramer;

hxy=0;
for q=1:factorialm
    for t=1:factorialm
        ni=ni+frec_obs(q,t);
        nj=nj+frec_obs(t,q);
        if frec_obs(t,q)~=0 %MODIFICADO MCRUIZ
            hxy=hxy-
(frec_obs(t,q)/ncramer)*log(frec_obs(t,q)/ncramer);
        end
    end

    tabla(10,q)=ni;
    tabla(11,q)=nj;
    ni=0;
    nj=0;
end

hx=0;
hy=0;
for q=1:factorialm
    if tabla(10,q)~=0 %MODIFICADO MCRUIZ
        hx=hx-(tabla(10,q)/ncramer)*log(tabla(10,q)/ncramer);
    end
    if tabla(11,q)~=0 %MODIFICADO MCRUIZ
        hy=hy-(tabla(11,q)/ncramer)*log(tabla(11,q)/ncramer);
    end
end

ixy=hx+hy-hxy;
d1=1-(ixy/hxy);

if hx>=hy
    d2=1-(ixy/hx);
else
    d2=1-(ixy/hy);
end
% dixy=1-ixy; %MODIFICADO MCRUIZ
dixy=1-(2*ixy/(hx+hy)); %MODIFICADO MCRUIZ

gtestfinal=gtest*2;
[RH01,PVAL1]=corr(serie1aux,serie2aux,'type','Spearman'); %metido
por spearman
[RH02,PVAL2]=corr(serie1aux,serie2aux,'type','Pearson');

crfinal=cr*2/(landa*(landa+1));

tabla(1,i)=suma_chi;
```




```

tabla(2,i)=crfinal;
tabla(3,i)=gtestfinal;
tabla(4,i)=37.65;
tabla(5,i)=583.61;
tabla(6,i)=0.05;
tabla(7,i)=m;
tabla(8,i)=PVAL1; %metido por spearman
tabla(9,i)=PVAL2;
tabla(12,i)=dvcramer;
tabla(13,i)=dixy;
tabla(14,i)=d1;
tabla(15,i)=d2;

```

end

end

%%%

b. Matrices de distancia

2005

Coef de Incertid.	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,82	0,88	0,70	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,82	0,00	0,96	0,91	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,88	0,96	0,00	0,83	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,70	0,91	0,83	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	1,00	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	1,00	0,00	0,99	0,91	0,90	1,00	1,00	0,94	0,81	0,90
Din. O	1,00	1,00	1,00	1,00	0,99	0,99	0,00	0,99	0,99	1,00	1,00	0,99	0,98	0,99
Finlandia	1,00	1,00	1,00	1,00	1,00	0,91	0,99	0,00	0,85	1,00	1,00	0,93	0,67	0,85
Nordpool	1,00	1,00	1,00	1,00	1,00	0,90	0,99	0,85	0,00	1,00	1,00	0,64	0,72	0,76
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,94	0,99	0,93	0,64	1,00	1,00	0,00	0,85	0,86
Suecia	1,00	1,00	1,00	1,00	1,00	0,81	0,98	0,67	0,72	1,00	1,00	0,85	0,00	0,72
Tro	1,00	1,00	1,00	1,00	1,00	0,90	0,99	0,85	0,76	1,00	1,00	0,86	0,72	0,00



Estudio de Dependencias en los Mercados Eléctricos



Universal	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,90	0,94	0,82	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,90	0,00	0,98	0,95	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,94	0,98	0,00	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,82	0,95	0,90	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	1,00	0,00	0,99	0,95	0,95	1,00	1,00	0,97	0,89	0,95
Din. O	1,00	1,00	1,00	1,00	1,00	0,99	0,00	0,99	0,99	1,00	1,00	0,99	0,99	0,99
Finlandia	1,00	1,00	1,00	1,00	1,00	0,95	0,99	0,00	0,92	1,00	1,00	0,96	0,80	0,92
Nordpool	1,00	1,00	1,00	1,00	1,00	0,95	0,99	0,92	0,00	1,00	1,00	0,78	0,84	0,87
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,97	0,99	0,96	0,78	1,00	1,00	0,00	0,92	0,93
Suecia	1,00	1,00	1,00	1,00	1,00	0,89	0,99	0,80	0,84	1,00	1,00	0,92	0,00	0,84
Tro	1,00	1,00	1,00	1,00	1,00	0,95	0,99	0,92	0,87	1,00	1,00	0,93	0,84	0,00

Universal 2	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,82	0,88	0,70	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,82	0,00	0,96	0,91	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,88	0,96	0,00	0,83	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,70	0,91	0,83	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	1,00	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	1,00	0,00	0,99	0,91	0,90	1,00	1,00	0,94	0,81	0,90
Din. O	1,00	1,00	1,00	1,00	0,99	0,99	0,00	0,99	0,99	1,00	1,00	0,99	0,99	0,99
Finlandia	1,00	1,00	1,00	1,00	1,00	0,91	0,99	0,00	0,85	1,00	1,00	0,93	0,67	0,85
Nordpool	1,00	1,00	1,00	1,00	1,00	0,90	0,99	0,85	0,00	1,00	1,00	0,64	0,72	0,77
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,94	0,99	0,93	0,64	1,00	1,00	0,00	0,85	0,86
Suecia	1,00	1,00	1,00	1,00	1,00	0,81	0,99	0,67	0,72	1,00	1,00	0,85	0,00	0,72
Tro	1,00	1,00	1,00	1,00	1,00	0,90	0,99	0,85	0,77	1,00	1,00	0,86	0,72	0,00



2006

Coef de Incertid.	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,85	0,85	0,70	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,85	0,00	0,96	0,93	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,85	0,96	0,00	0,74	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,70	0,93	0,74	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	1,00	0,00	0,97	0,87	0,94	1,00	1,00	0,99	0,84	0,95
Din. O	1,00	1,00	1,00	1,00	1,00	0,97	0,00	0,98	0,97	1,00	1,00	0,99	0,97	0,99
Finlandia	1,00	1,00	1,00	1,00	1,00	0,87	0,98	0,00	0,88	1,00	1,00	0,97	0,52	0,84
Nordpool	1,00	1,00	1,00	1,00	1,00	0,94	0,97	0,88	0,00	1,00	1,00	0,85	0,83	0,88
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,99	0,99	0,97	0,85	1,00	1,00	0,00	0,93	0,94
Suecia	1,00	1,00	1,00	1,00	1,00	0,84	0,97	0,52	0,83	1,00	1,00	0,93	0,00	0,70
Tro	1,00	1,00	1,00	1,00	1,00	0,95	0,99	0,84	0,88	1,00	1,00	0,94	0,70	0,00

Universal	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,92	0,92	0,82	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,92	0,00	0,98	0,96	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,92	0,98	0,00	0,85	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,82	0,96	0,85	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	1,00	0,00	0,98	0,93	0,97	1,00	1,00	0,99	0,91	0,97
Din. O	1,00	1,00	1,00	1,00	1,00	0,98	0,00	0,99	0,98	1,00	1,00	1,00	0,99	1,00
Finlandia	1,00	1,00	1,00	1,00	1,00	0,93	0,99	0,00	0,94	1,00	1,00	0,98	0,69	0,91
Nordpool	1,00	1,00	1,00	1,00	1,00	0,97	0,98	0,94	0,00	1,00	1,00	0,92	0,91	0,94
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,99	1,00	0,98	0,92	1,00	1,00	0,00	0,97	0,97
Suecia	1,00	1,00	1,00	1,00	1,00	0,91	0,99	0,69	0,91	1,00	1,00	0,97	0,00	0,83
Tro	1,00	1,00	1,00	1,00	1,00	0,97	1,00	0,91	0,94	1,00	1,00	0,97	0,83	0,00



<u>Universal</u> <u>2</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,85	0,86	0,70	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,85	0,00	0,96	0,93	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,86	0,96	0,00	0,75	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,70	0,93	0,75	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	1,00	0,00	0,97	0,87	0,94	1,00	1,00	0,99	0,84	0,95
Din. O	1,00	1,00	1,00	1,00	1,00	0,97	0,00	0,98	0,97	1,00	1,00	0,99	0,97	0,99
Finlandia	1,00	1,00	1,00	1,00	1,00	0,87	0,98	0,00	0,88	1,00	1,00	0,97	0,52	0,84
Nordpool	1,00	1,00	1,00	1,00	1,00	0,94	0,97	0,88	0,00	1,00	1,00	0,85	0,83	0,88
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,99	0,99	0,97	0,85	1,00	1,00	0,00	0,93	0,94
Suecia	1,00	1,00	1,00	1,00	1,00	0,84	0,97	0,52	0,83	1,00	1,00	0,93	0,00	0,71
Tro	1,00	1,00	1,00	1,00	1,00	0,95	0,99	0,84	0,88	1,00	1,00	0,94	0,71	0,00

2007

<u>Coef de</u> <u>Incertid.</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,69	0,83	0,71	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,69	0,00	0,92	0,86	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,83	0,92	0,00	0,65	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,71	0,86	0,65	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	0,99	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	0,99	0,00	0,93	0,91	0,95	1,00	1,00	0,99	0,89	0,95
Din. O	1,00	1,00	1,00	1,00	0,99	0,93	0,00	0,99	0,97	1,00	1,00	1,00	0,98	0,99
Finlandia	1,00	1,00	1,00	1,00	1,00	0,91	0,99	0,00	0,87	1,00	1,00	0,96	0,36	0,75
Nordpool	1,00	1,00	1,00	1,00	1,00	0,95	0,97	0,87	0,00	1,00	1,00	0,89	0,84	0,87
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,99	1,00	0,96	0,89	1,00	1,00	0,00	0,96	0,95
Suecia	1,00	1,00	1,00	1,00	1,00	0,89	0,98	0,36	0,84	1,00	1,00	0,96	0,00	0,67
Tro	1,00	1,00	1,00	1,00	1,00	0,95	0,99	0,75	0,87	1,00	1,00	0,95	0,67	0,00



<u>Universal</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,82	0,91	0,83	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,82	0,00	0,96	0,92	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,91	0,96	0,00	0,79	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,83	0,92	0,79	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	1,00	0,00	0,96	0,95	0,97	1,00	1,00	0,99	0,94	0,97
Din. O	1,00	1,00	1,00	1,00	1,00	0,96	0,00	0,99	0,98	1,00	1,00	1,00	0,99	1,00
Finlandia	1,00	1,00	1,00	1,00	1,00	0,95	0,99	0,00	0,93	1,00	1,00	0,98	0,53	0,86
Nordpool	1,00	1,00	1,00	1,00	1,00	0,97	0,98	0,93	0,00	1,00	1,00	0,94	0,91	0,93
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,99	1,00	0,98	0,94	1,00	1,00	0,00	0,98	0,97
Suecia	1,00	1,00	1,00	1,00	1,00	0,94	0,99	0,53	0,91	1,00	1,00	0,98	0,00	0,80
Tro	1,00	1,00	1,00	1,00	1,00	0,97	1,00	0,86	0,93	1,00	1,00	0,97	0,80	0,00

<u>Universal 2</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,70	0,83	0,71	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,70	0,00	0,92	0,86	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,83	0,92	0,00	0,65	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,71	0,86	0,65	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	0,99	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	0,99	0,00	0,93	0,91	0,95	1,00	1,00	0,99	0,89	0,95
Din. O	1,00	1,00	1,00	1,00	0,99	0,93	0,00	0,99	0,97	1,00	1,00	1,00	0,98	0,99
Finlandia	1,00	1,00	1,00	1,00	1,00	0,91	0,99	0,00	0,87	1,00	1,00	0,96	0,36	0,75
Nordpool	1,00	1,00	1,00	1,00	1,00	0,95	0,97	0,87	0,00	1,00	1,00	0,89	0,84	0,87
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,99	1,00	0,96	0,89	1,00	1,00	0,00	0,96	0,95
Suecia	1,00	1,00	1,00	1,00	1,00	0,89	0,98	0,36	0,84	1,00	1,00	0,96	0,00	0,67
Tro	1,00	1,00	1,00	1,00	1,00	0,95	0,99	0,75	0,87	1,00	1,00	0,95	0,67	0,00



2008

<u>Coef de Incertid.</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,76	0,86	0,77	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,76	0,00	0,93	0,89	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,86	0,93	0,00	0,70	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,77	0,89	0,70	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	0,99	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	0,99	0,00	0,93	0,91	0,97	1,00	1,00	1,00	0,90	0,97
Din. O	1,00	1,00	1,00	1,00	0,99	0,93	0,00	0,97	0,98	1,00	1,00	1,00	0,97	0,99
Finlandia	1,00	1,00	1,00	1,00	1,00	0,91	0,97	0,00	0,90	1,00	1,00	1,00	0,30	0,87
Nordpool	1,00	1,00	1,00	1,00	1,00	0,97	0,98	0,90	0,00	1,00	1,00	0,95	0,88	0,93
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,00	0,99	1,00
Suecia	1,00	1,00	1,00	1,00	1,00	0,90	0,97	0,30	0,88	1,00	1,00	0,99	0,00	0,83
Tro	1,00	1,00	1,00	1,00	1,00	0,97	0,99	0,87	0,93	1,00	1,00	1,00	0,83	0,00

<u>Universal</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,86	0,92	0,87	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,86	0,00	0,96	0,94	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,92	0,96	0,00	0,82	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,87	0,94	0,82	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	1,00	0,00	0,97	0,95	0,98	1,00	1,00	1,00	0,95	0,99
Din. O	1,00	1,00	1,00	1,00	1,00	0,97	0,00	0,99	0,99	1,00	1,00	1,00	0,98	1,00
Finlandia	1,00	1,00	1,00	1,00	1,00	0,95	0,99	0,00	0,95	1,00	1,00	1,00	0,46	0,93
Nordpool	1,00	1,00	1,00	1,00	1,00	0,98	0,99	0,95	0,00	1,00	1,00	0,98	0,94	0,97
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,98	1,00	1,00	0,00	1,00	1,00
Suecia	1,00	1,00	1,00	1,00	1,00	0,95	0,98	0,46	0,94	1,00	1,00	1,00	0,00	0,91
Tro	1,00	1,00	1,00	1,00	1,00	0,99	1,00	0,93	0,97	1,00	1,00	1,00	0,91	0,00



<u>Universal</u> <u>2</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,76	0,86	0,77	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,76	0,00	0,93	0,89	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,86	0,93	0,00	0,70	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,77	0,89	0,70	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	0,99	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	0,99	0,00	0,93	0,91	0,97	1,00	1,00	1,00	0,90	0,97
Din. O	1,00	1,00	1,00	1,00	0,99	0,93	0,00	0,97	0,98	1,00	1,00	1,00	0,97	0,99
Finlandia	1,00	1,00	1,00	1,00	1,00	0,91	0,97	0,00	0,90	1,00	1,00	1,00	0,30	0,87
Nordpool	1,00	1,00	1,00	1,00	1,00	0,97	0,98	0,90	0,00	1,00	1,00	0,95	0,88	0,93
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,95	1,00	1,00	0,00	0,99	1,00
Suecia	1,00	1,00	1,00	1,00	1,00	0,90	0,97	0,30	0,88	1,00	1,00	0,99	0,00	0,83
Tro	1,00	1,00	1,00	1,00	1,00	0,97	0,99	0,87	0,93	1,00	1,00	1,00	0,83	0,00

2009

<u>Coef de</u> <u>Incertid.</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,71	0,89	0,82	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,71	0,00	0,95	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,89	0,95	0,00	0,77	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,82	0,90	0,77	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	0,99	0,99	0,99	0,99	1,00	1,00	1,00	0,99	1,00
Din. E	1,00	1,00	1,00	1,00	0,99	0,00	0,95	0,85	0,90	1,00	1,00	0,96	0,80	0,94
Din. O	1,00	1,00	1,00	1,00	0,99	0,95	0,00	0,95	0,93	1,00	1,00	0,98	0,95	0,98
Finlandia	1,00	1,00	1,00	1,00	0,99	0,85	0,95	0,00	0,83	1,00	1,00	0,96	0,34	0,82
Nordpool	1,00	1,00	1,00	1,00	0,99	0,90	0,93	0,83	0,00	1,00	1,00	0,84	0,81	0,87
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,96	0,98	0,96	0,84	1,00	1,00	0,00	0,95	0,96
Suecia	1,00	1,00	1,00	1,00	0,99	0,80	0,95	0,34	0,81	1,00	1,00	0,95	0,00	0,76
Tro	1,00	1,00	1,00	1,00	1,00	0,94	0,98	0,82	0,87	1,00	1,00	0,96	0,76	0,00



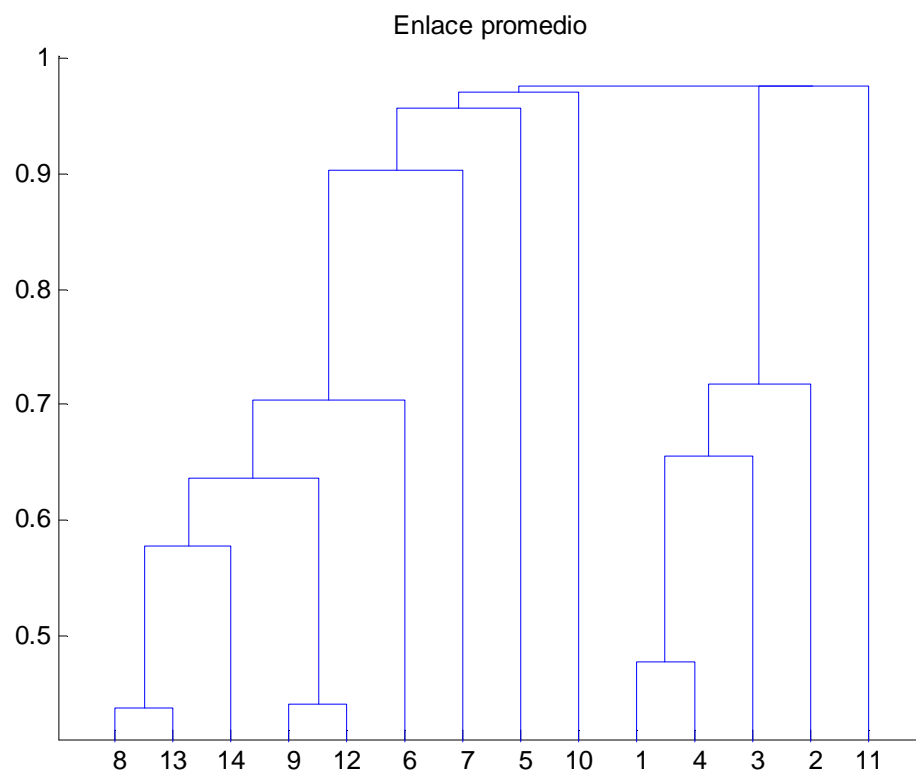
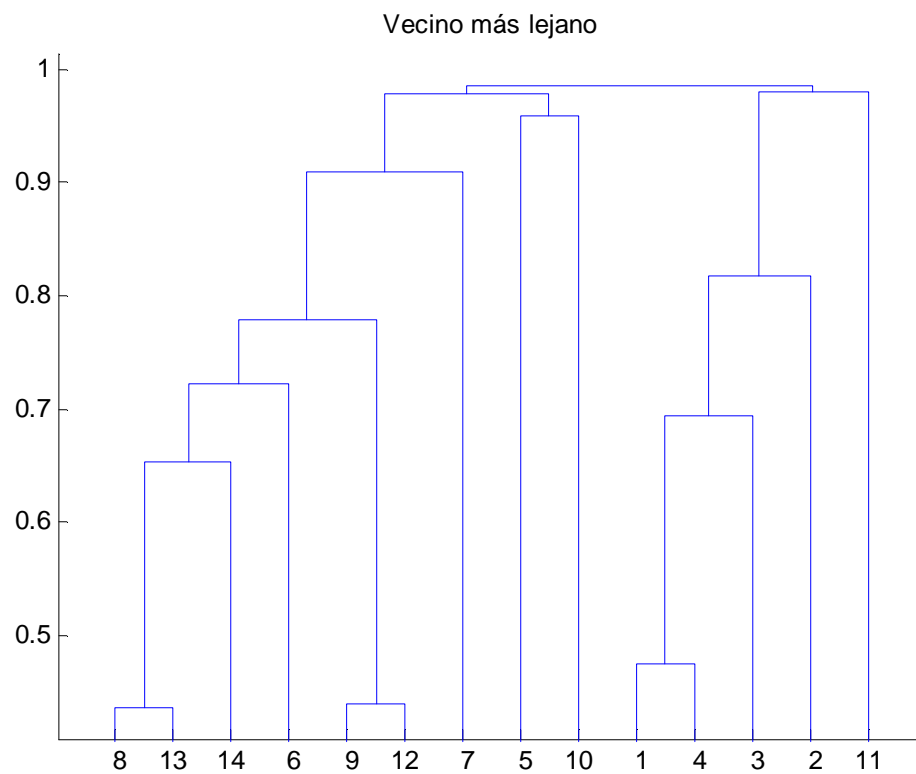
<u>Universal</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,83	0,94	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,83	0,00	0,97	0,95	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,94	0,97	0,00	0,87	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,90	0,95	0,87	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Din. E	1,00	1,00	1,00	1,00	0,99	0,00	0,97	0,92	0,95	1,00	1,00	0,98	0,89	0,97
Din. O	1,00	1,00	1,00	1,00	1,00	0,97	0,00	0,98	0,96	1,00	1,00	0,99	0,97	0,99
Finlandia	1,00	1,00	1,00	1,00	1,00	0,92	0,98	0,00	0,91	1,00	1,00	0,98	0,50	0,90
Nordpool	1,00	1,00	1,00	1,00	1,00	0,95	0,96	0,91	0,00	1,00	1,00	0,91	0,89	0,93
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,98	0,99	0,98	0,91	1,00	1,00	0,00	0,97	0,98
Suecia	1,00	1,00	1,00	1,00	1,00	0,89	0,97	0,50	0,89	1,00	1,00	0,97	0,00	0,86
Tro	1,00	1,00	1,00	1,00	1,00	0,97	0,99	0,90	0,93	1,00	1,00	0,98	0,86	0,00

<u>Universal 2</u>	NSW	QLD	SA	VIC	Austria	Din. E	Din. O	Finl	Nord	OMEL	Ont	Oslo	Suec	Tro
NSW	0,00	0,71	0,89	0,83	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
QLD	0,71	0,00	0,95	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,89	0,95	0,00	0,77	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
VIC	0,83	0,90	0,77	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Austria	1,00	1,00	1,00	1,00	0,00	0,99	0,99	0,99	0,99	1,00	1,00	1,00	0,99	1,00
Din. E	1,00	1,00	1,00	1,00	0,99	0,00	0,95	0,85	0,90	1,00	1,00	0,96	0,80	0,94
Din. O	1,00	1,00	1,00	1,00	0,99	0,95	0,00	0,95	0,93	1,00	1,00	0,98	0,95	0,98
Finlandia	1,00	1,00	1,00	1,00	0,99	0,85	0,95	0,00	0,84	1,00	1,00	0,96	0,34	0,82
Nordpool	1,00	1,00	1,00	1,00	0,99	0,90	0,93	0,84	0,00	1,00	1,00	0,84	0,81	0,87
OMEL	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00
Ontario	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00
Oslo	1,00	1,00	1,00	1,00	1,00	0,96	0,98	0,96	0,84	1,00	1,00	0,00	0,95	0,96
Suecia	1,00	1,00	1,00	1,00	0,99	0,80	0,95	0,34	0,81	1,00	1,00	0,95	0,00	0,76
Tro	1,00	1,00	1,00	1,00	1,00	0,94	0,98	0,82	0,87	1,00	1,00	0,96	0,76	0,00

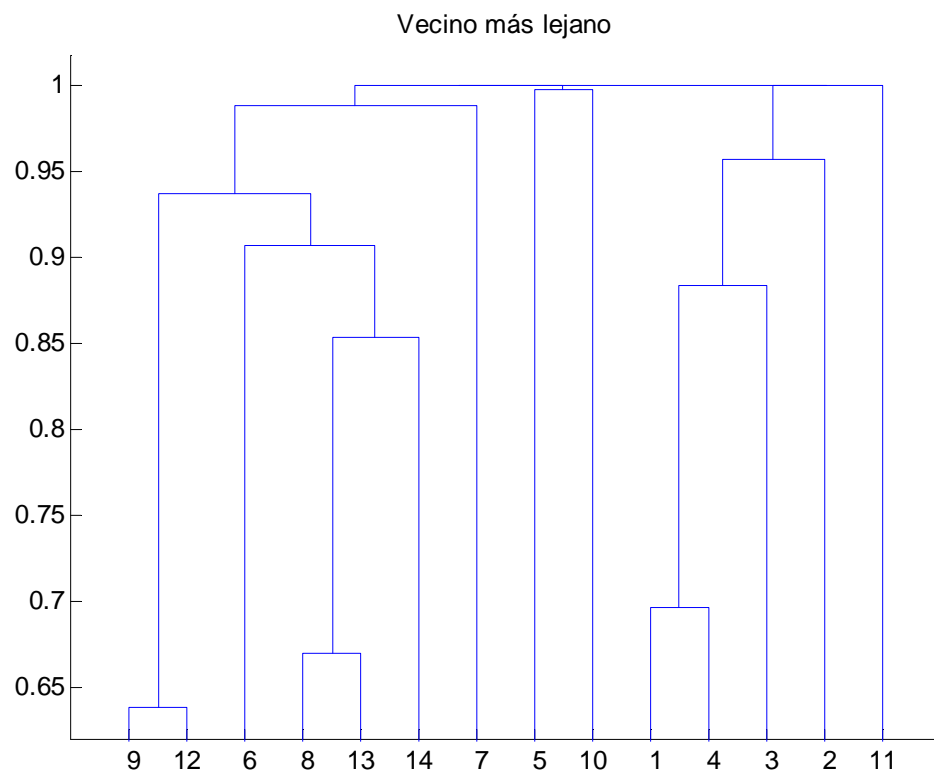
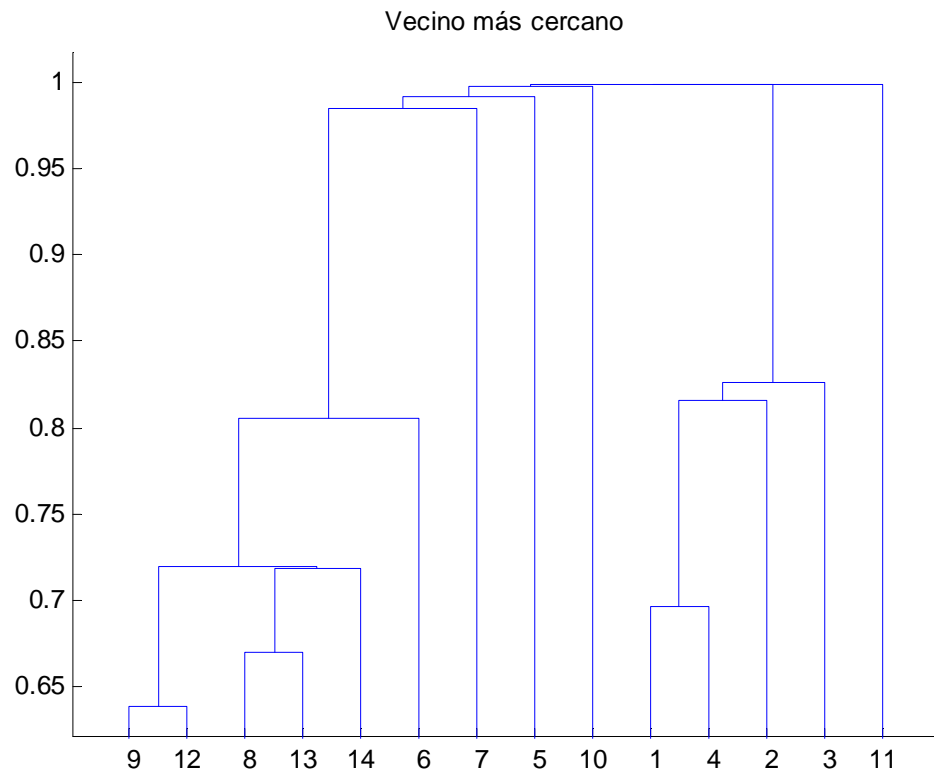


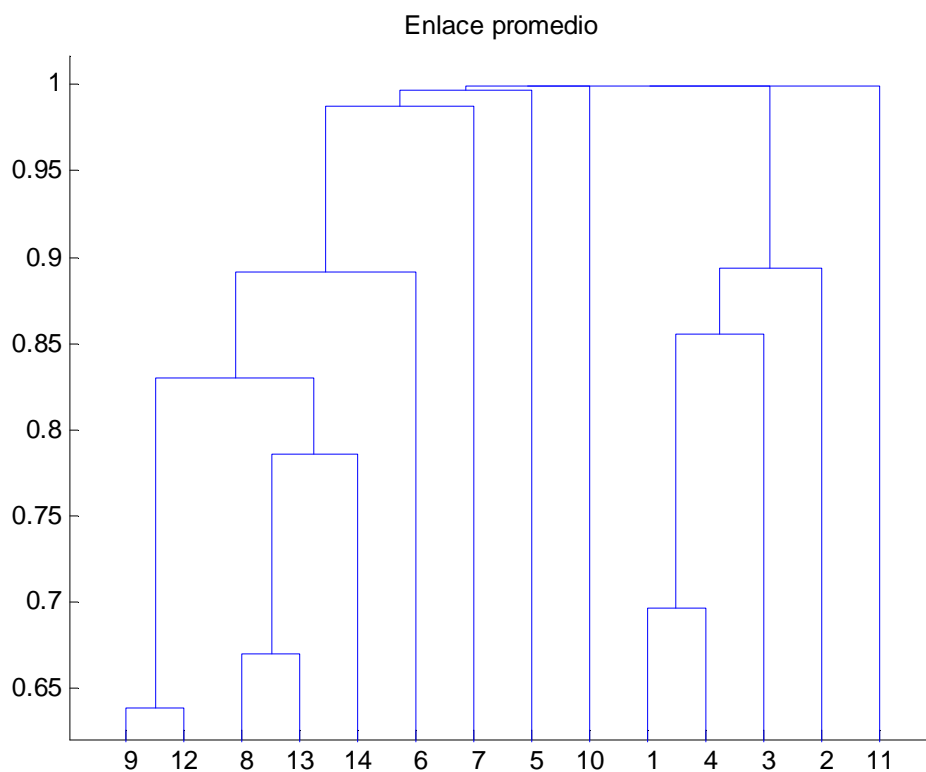
c. Dendogramas

- Cramer 2005

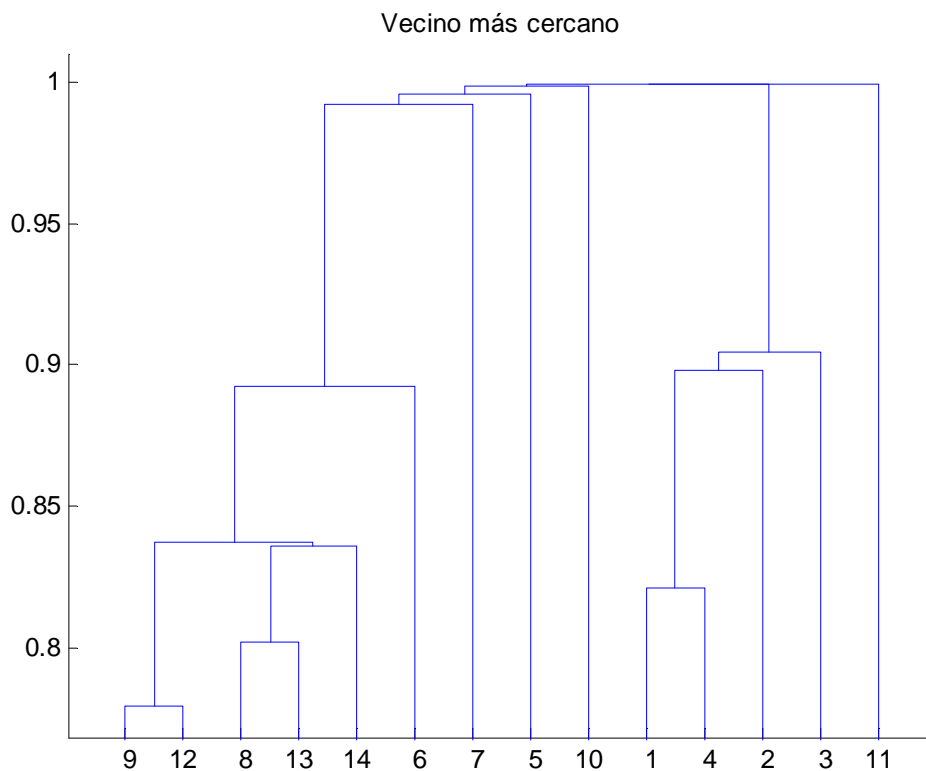


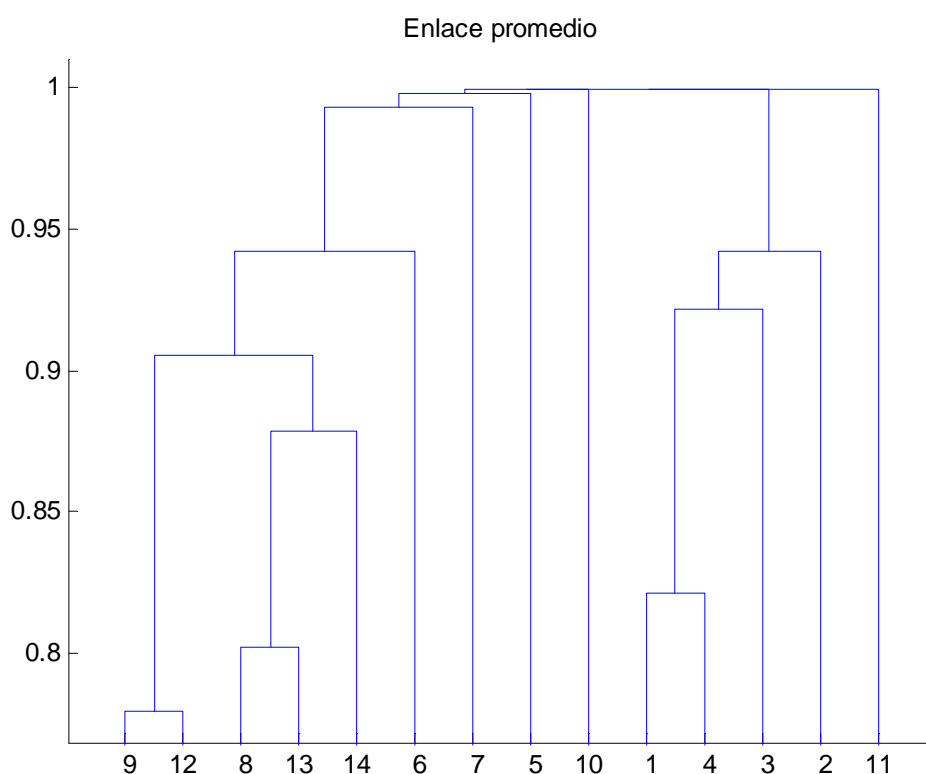
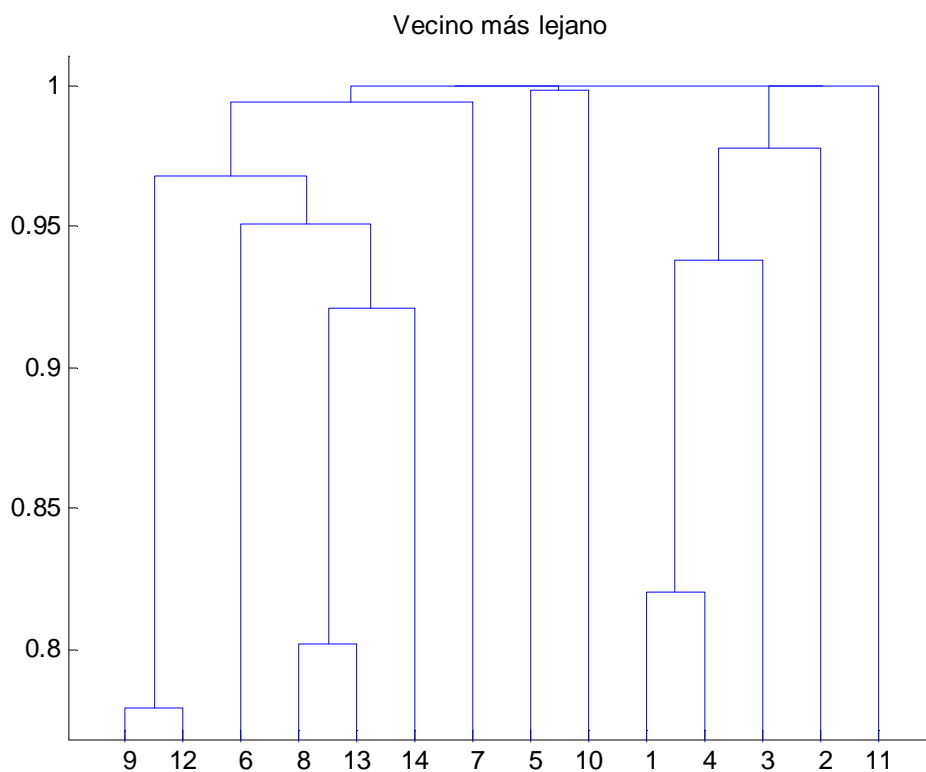
- DCI 2005





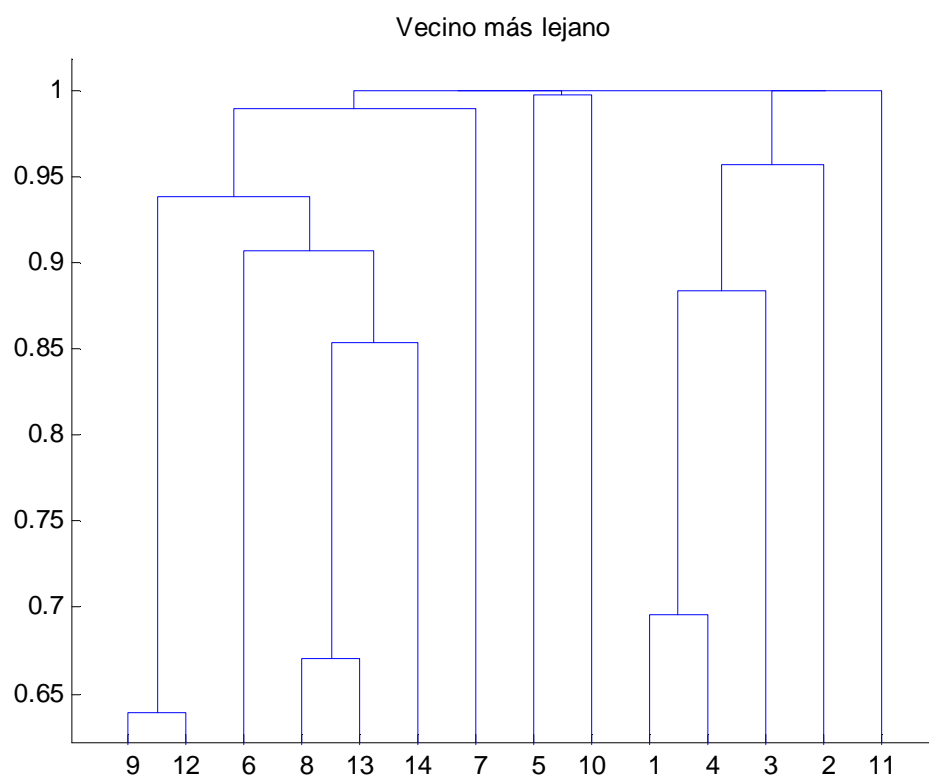
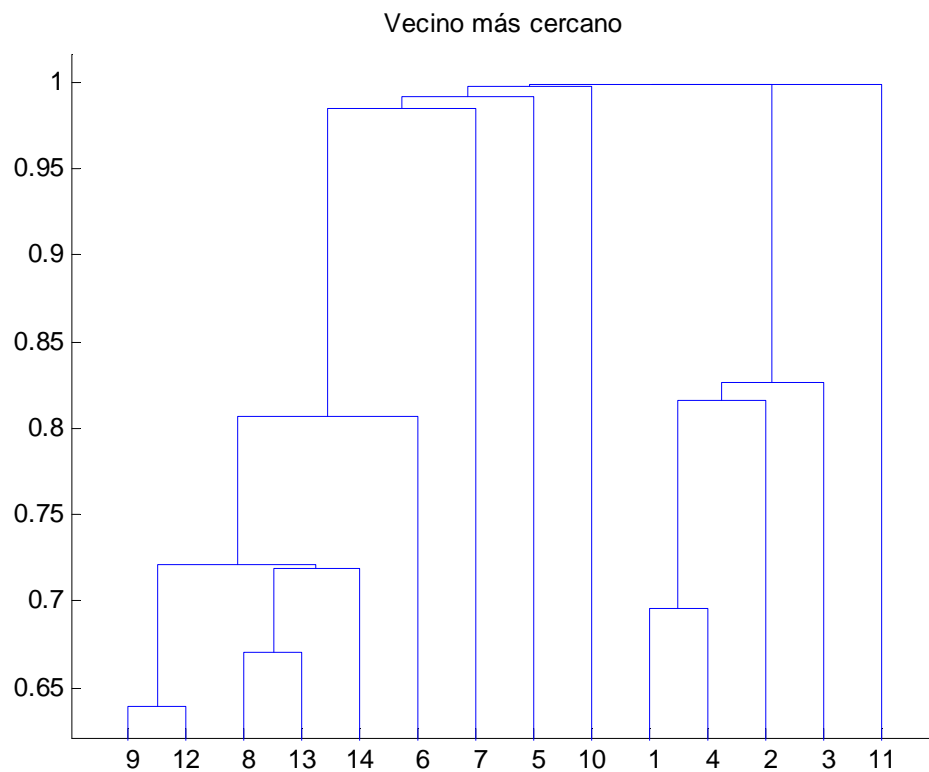
- DU1,2005

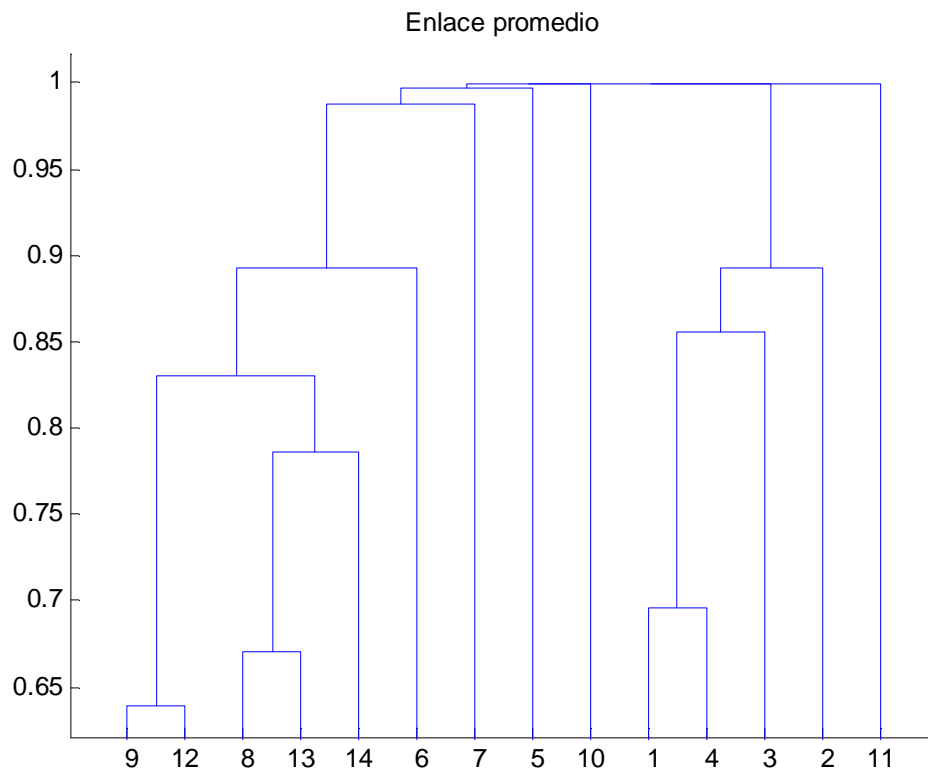




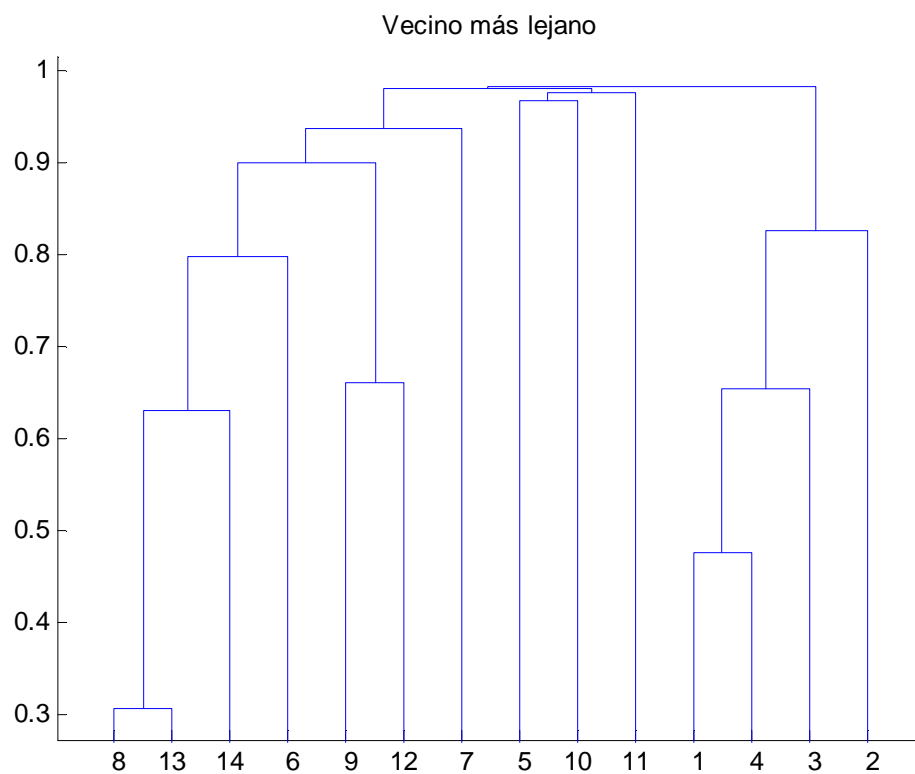


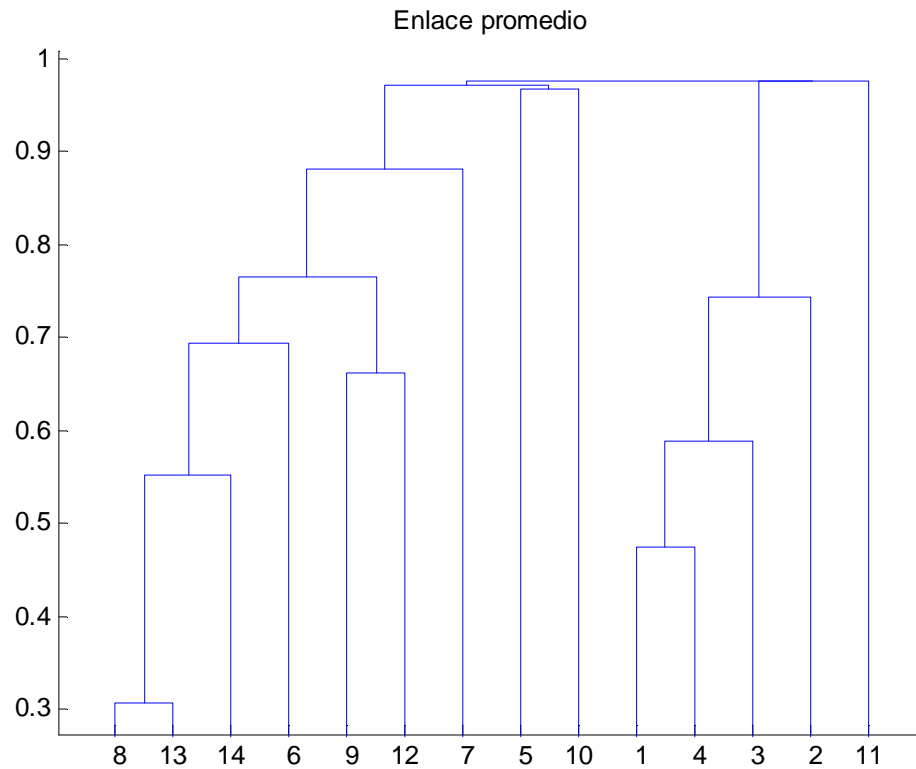
- DU2



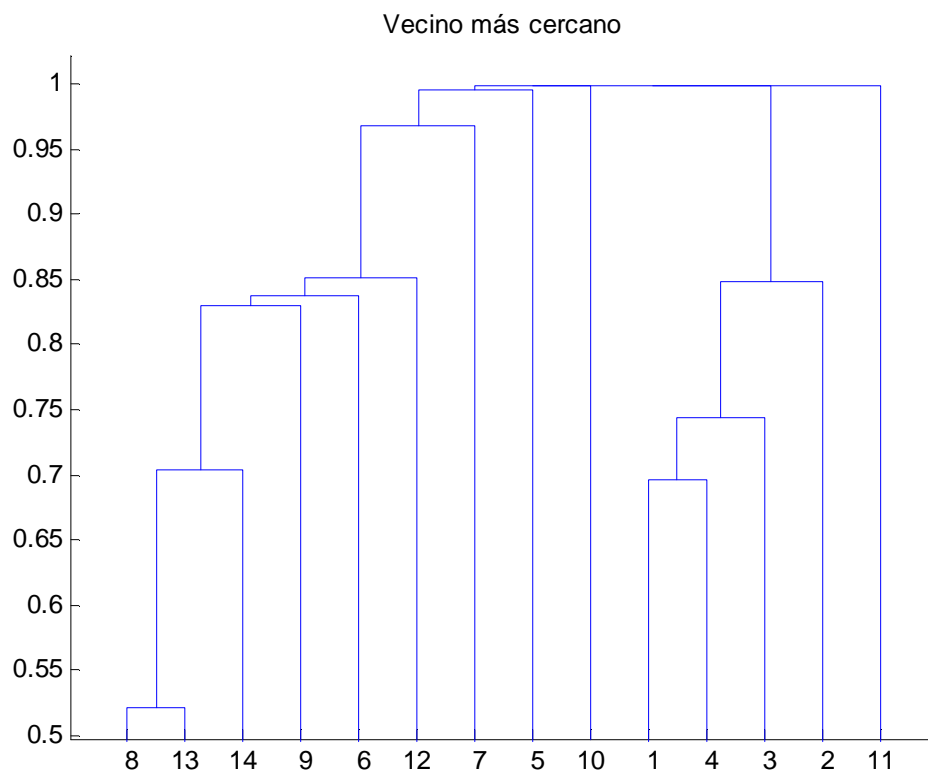


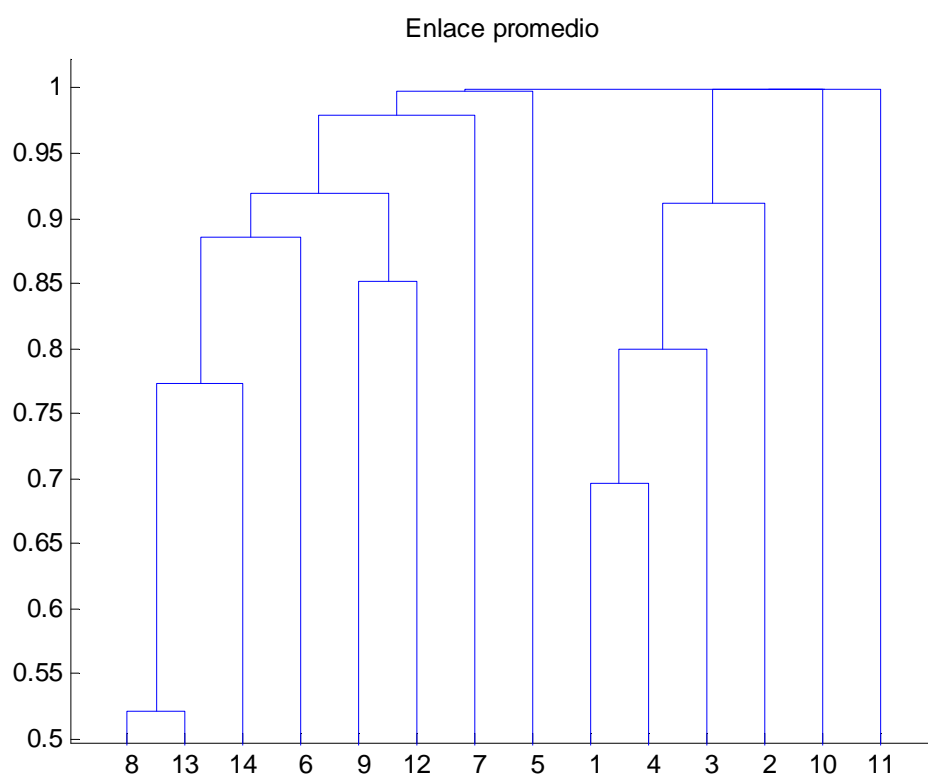
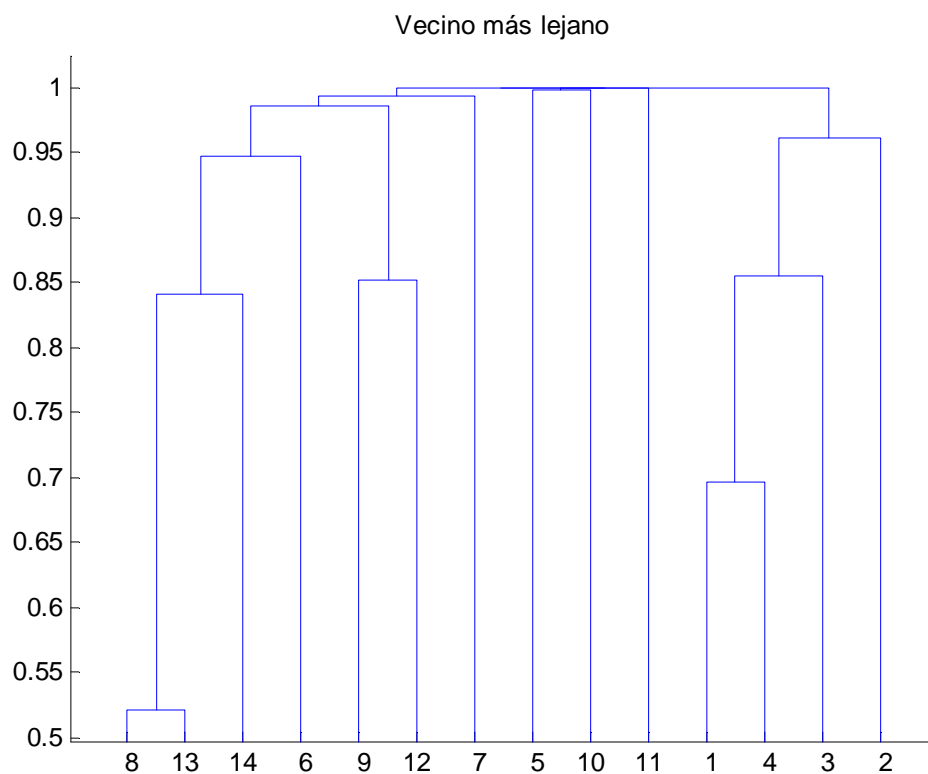
- Cramer, 2006





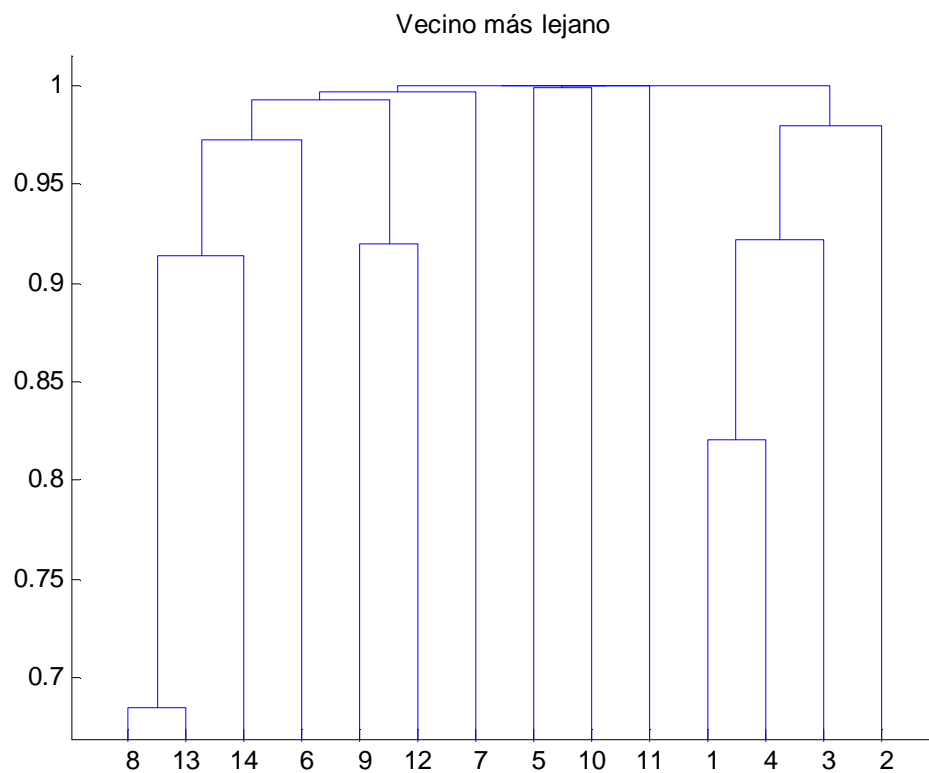
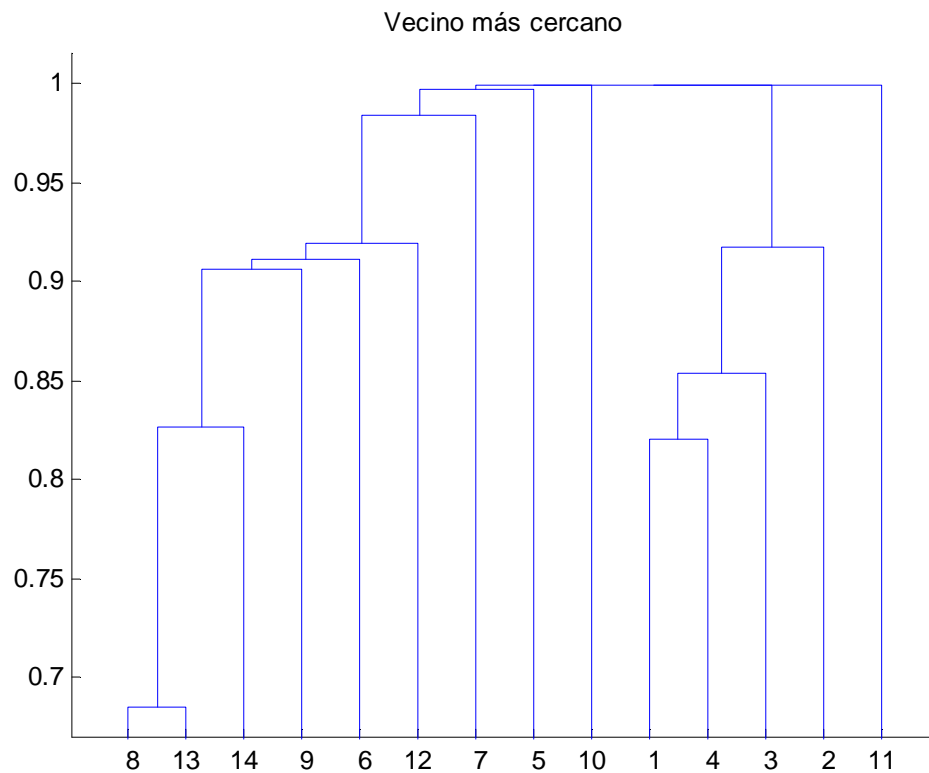
- DCI, 2006

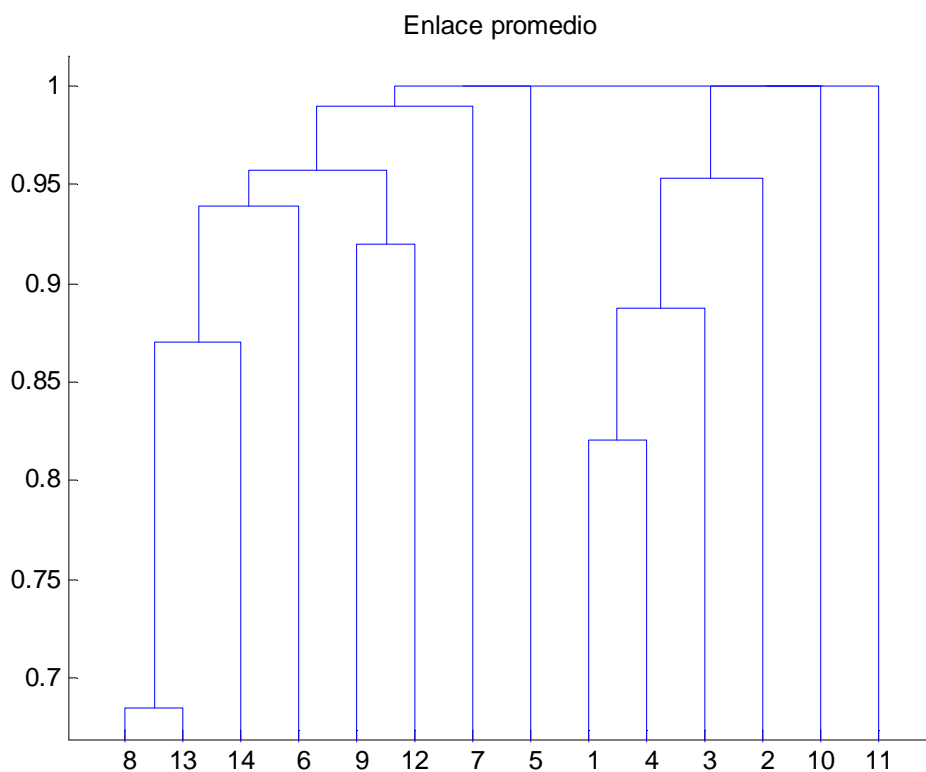




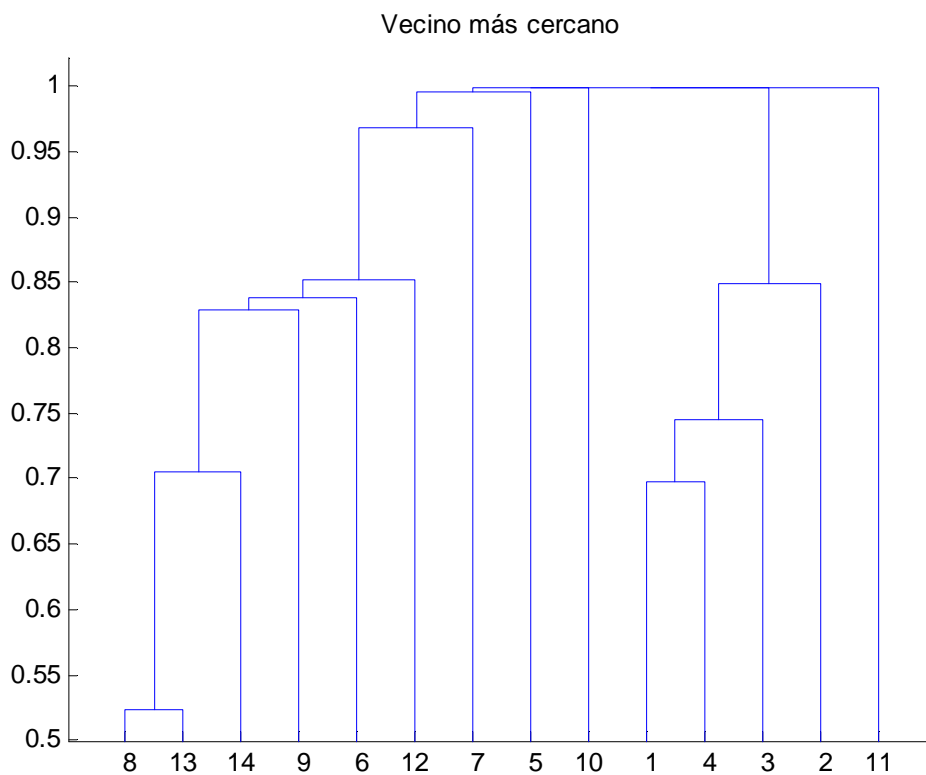


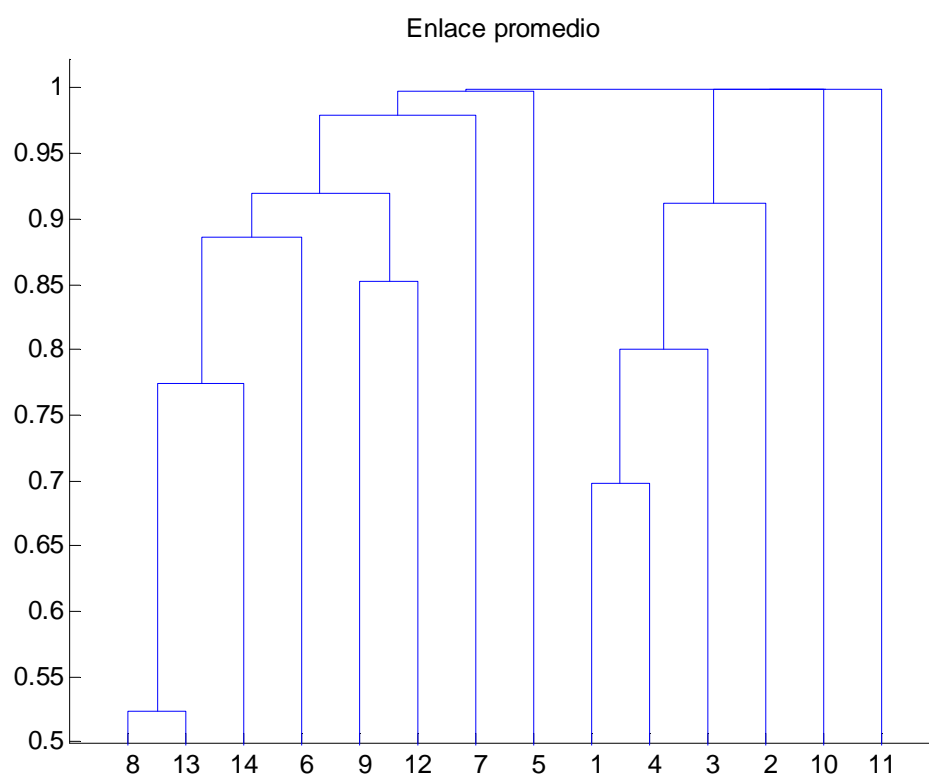
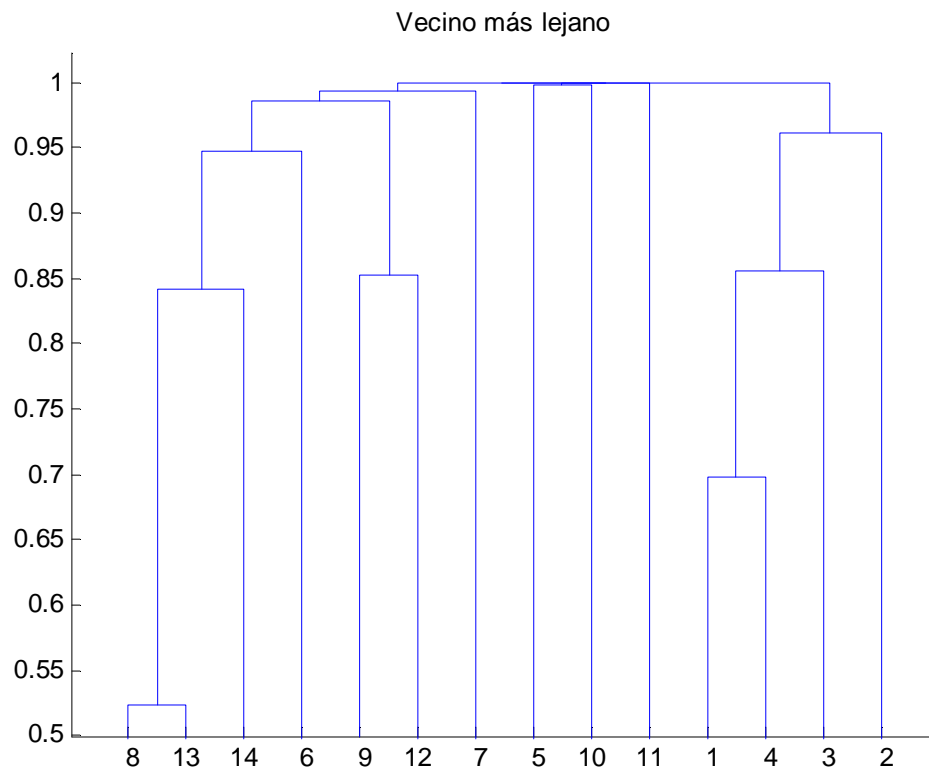
- DU1, 2006





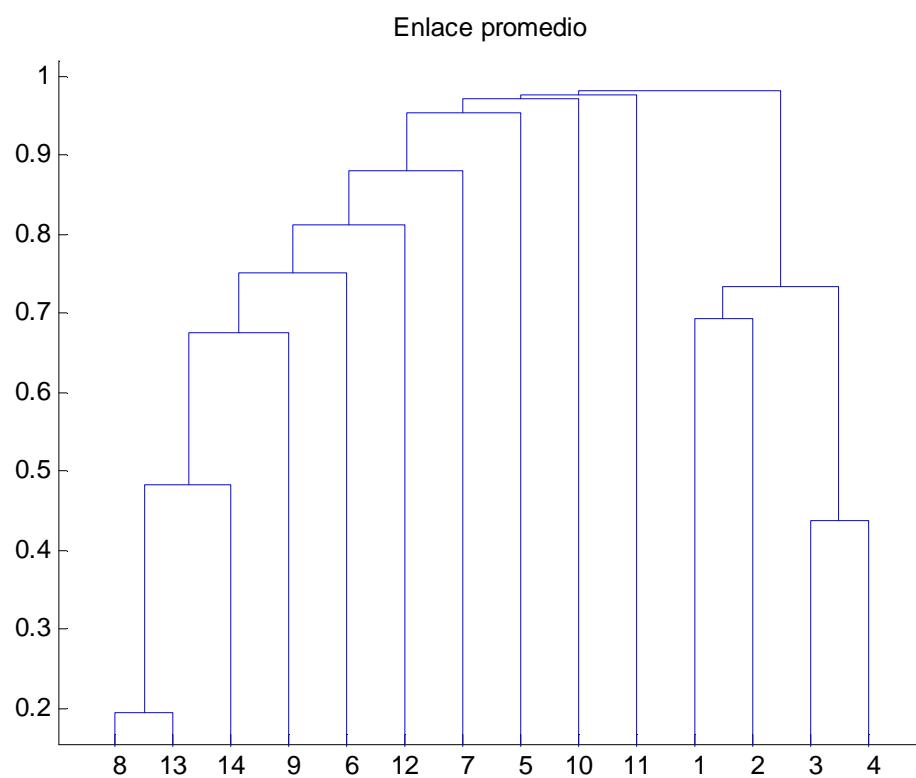
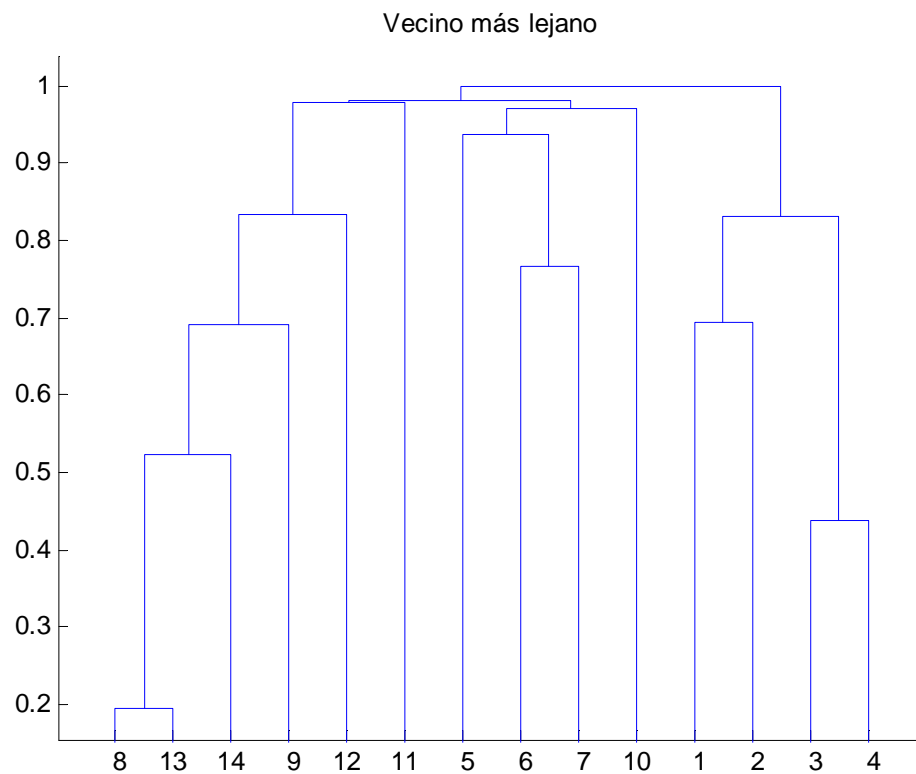
- DU2, 2006





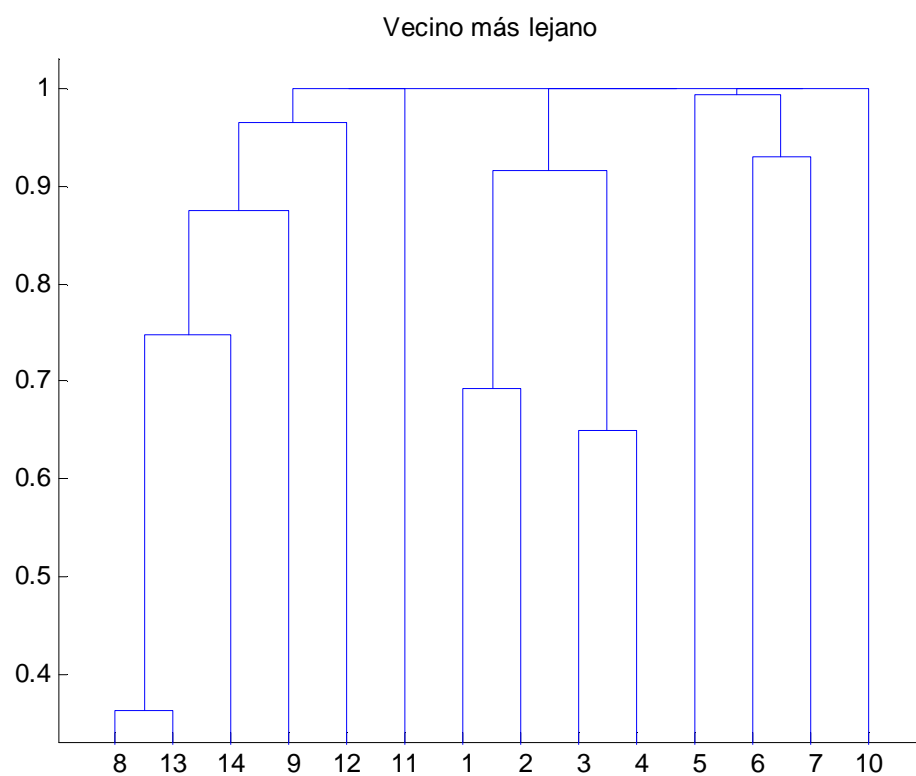
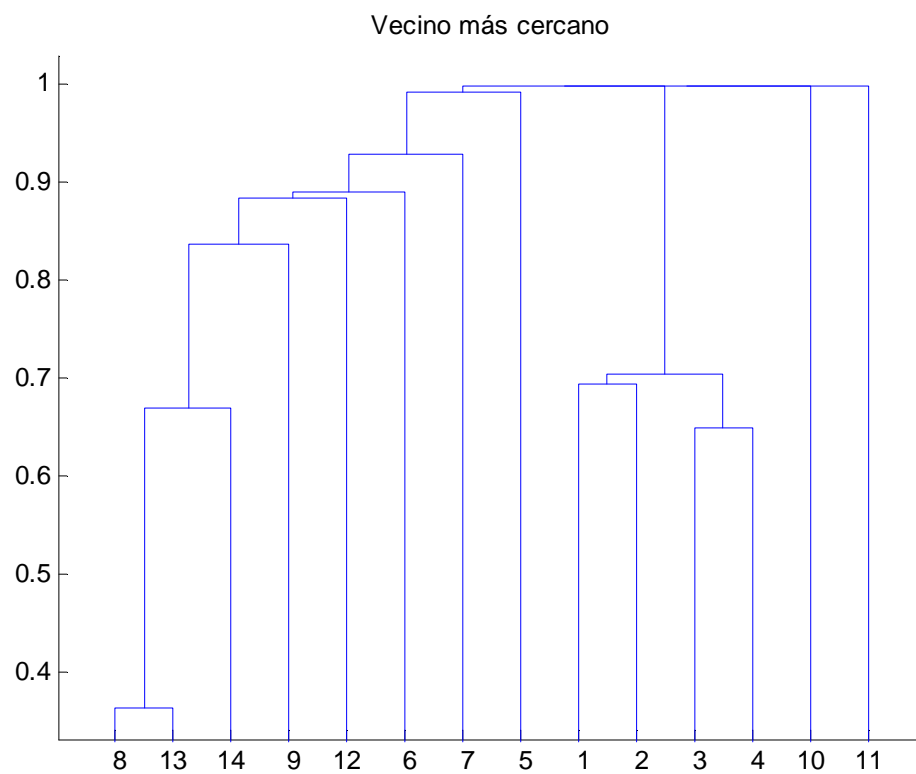


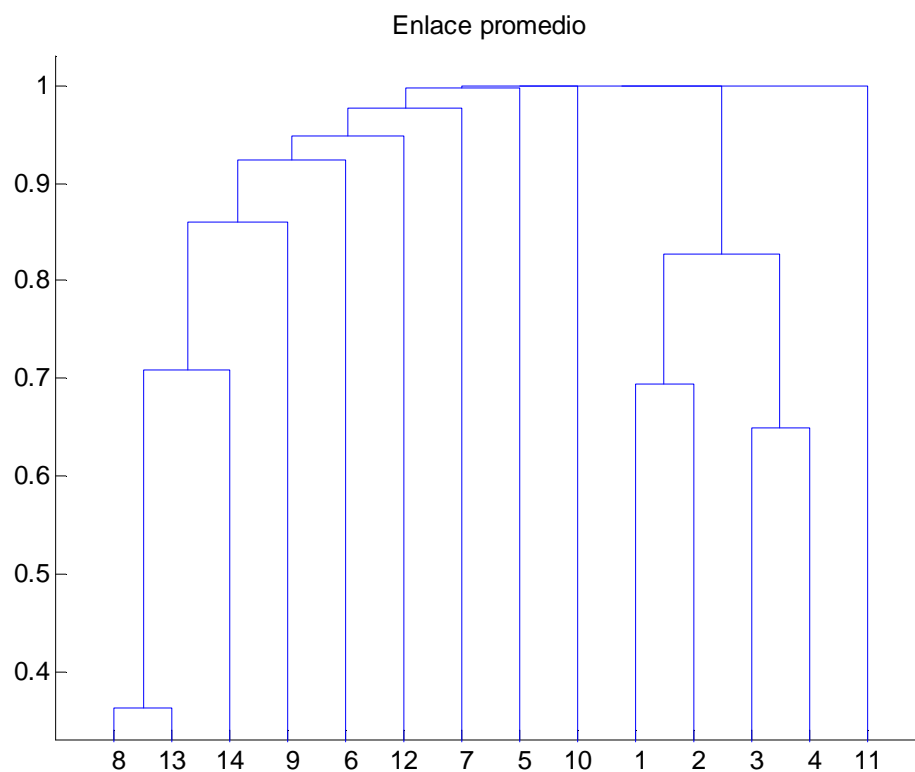
- Dcramer, 2007



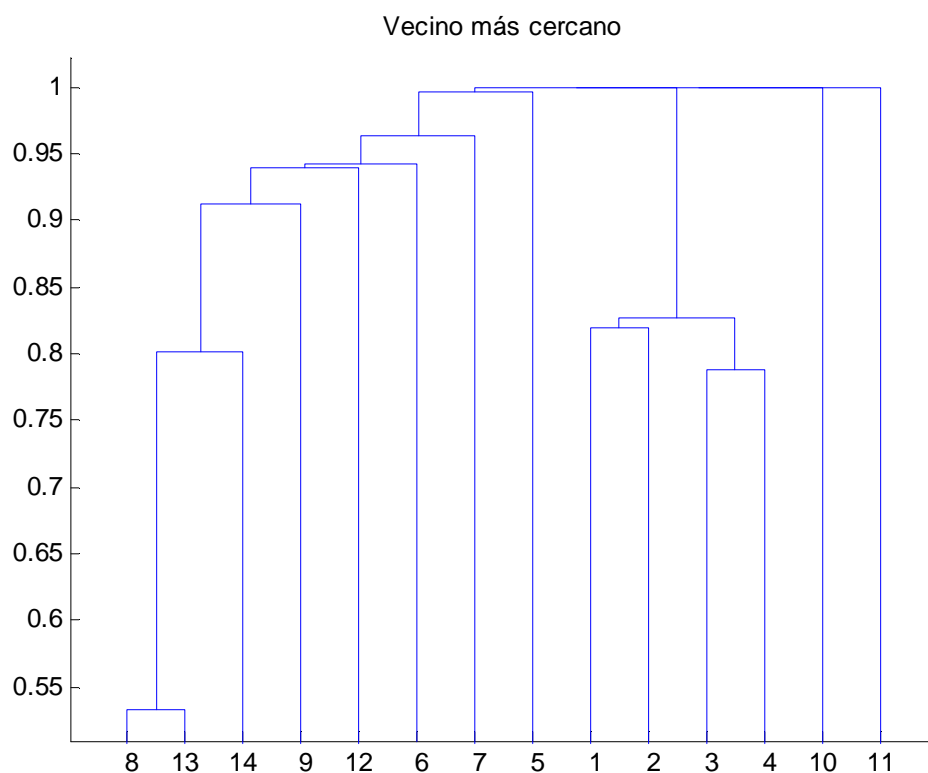


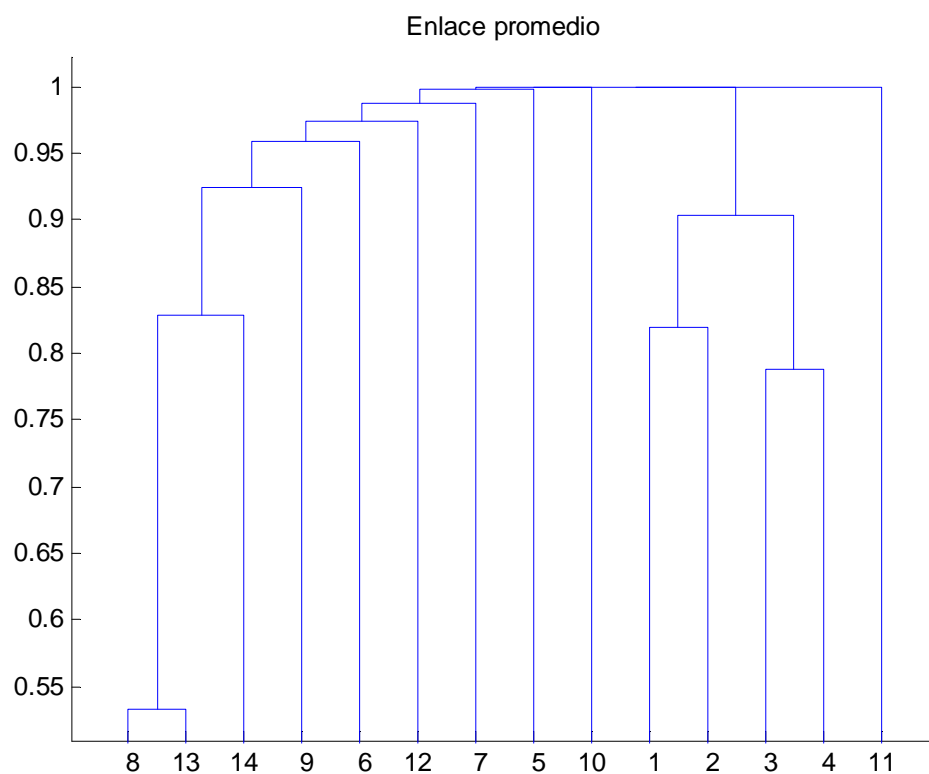
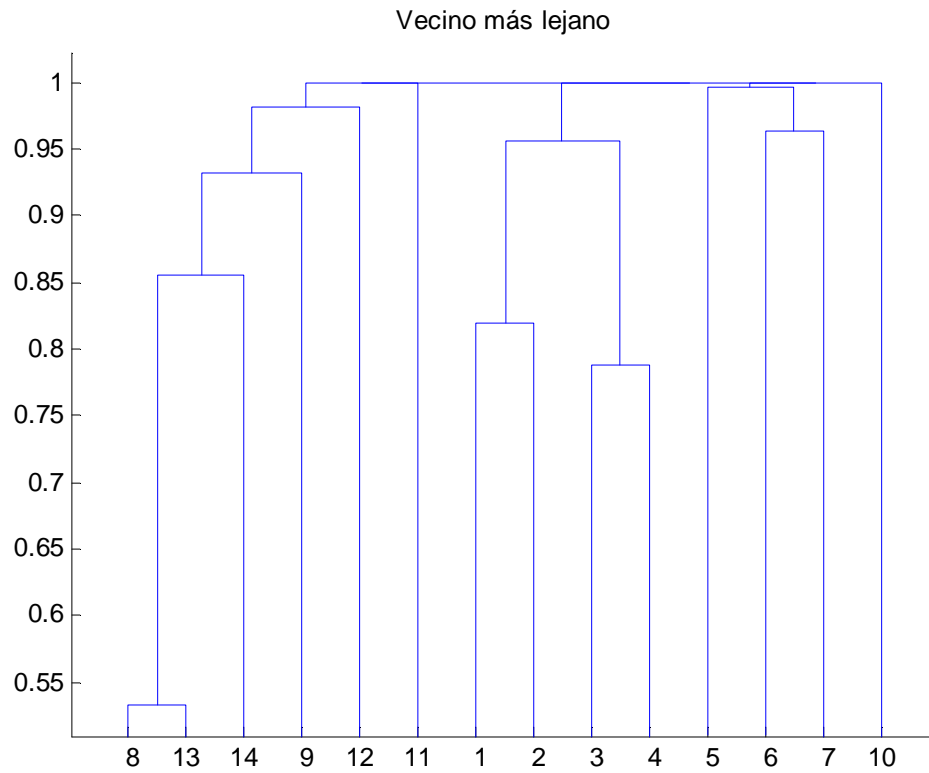
- DCI,2007





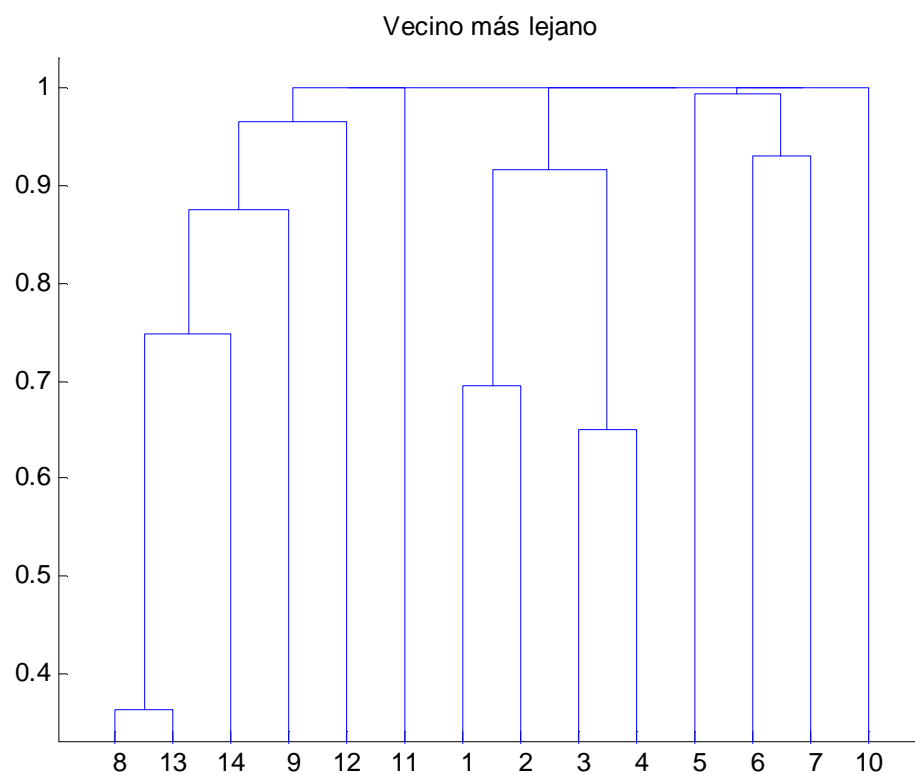
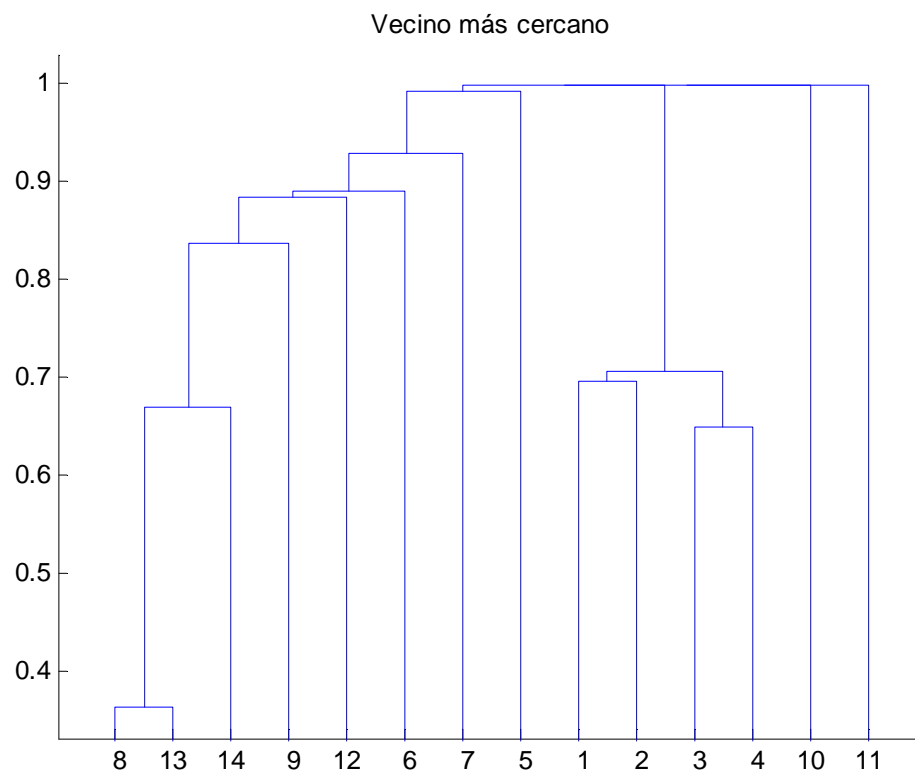
- DU1, 2007

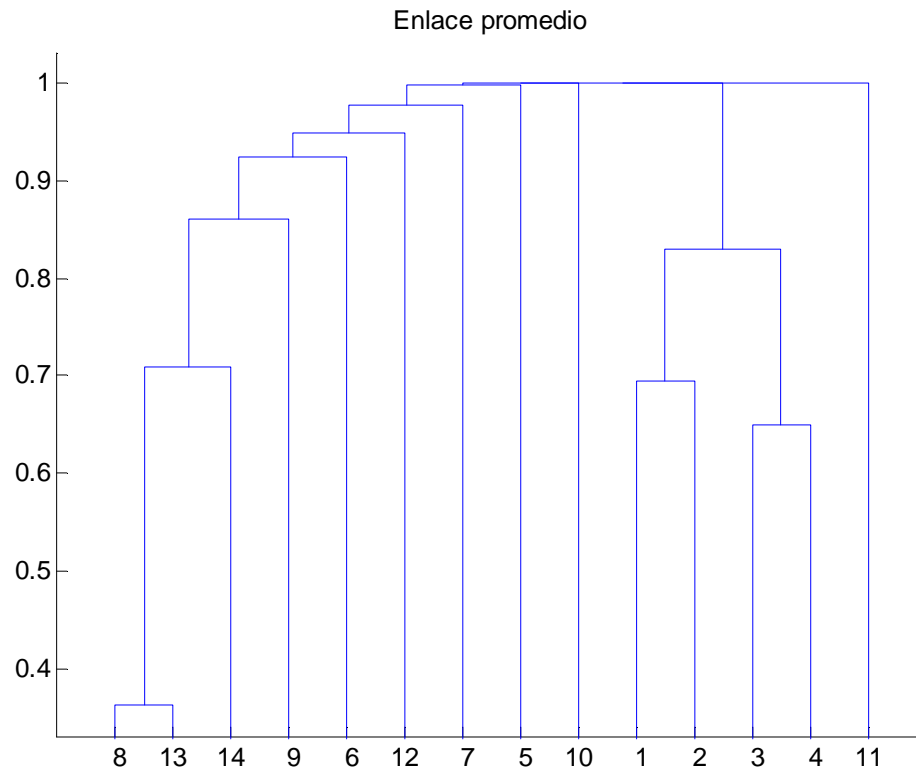




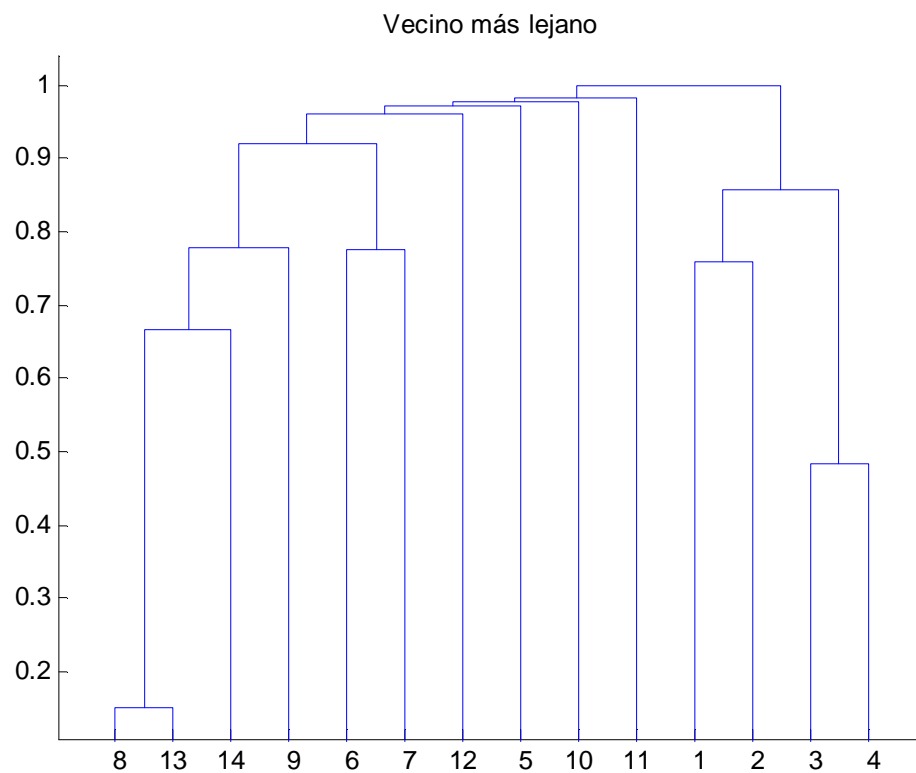


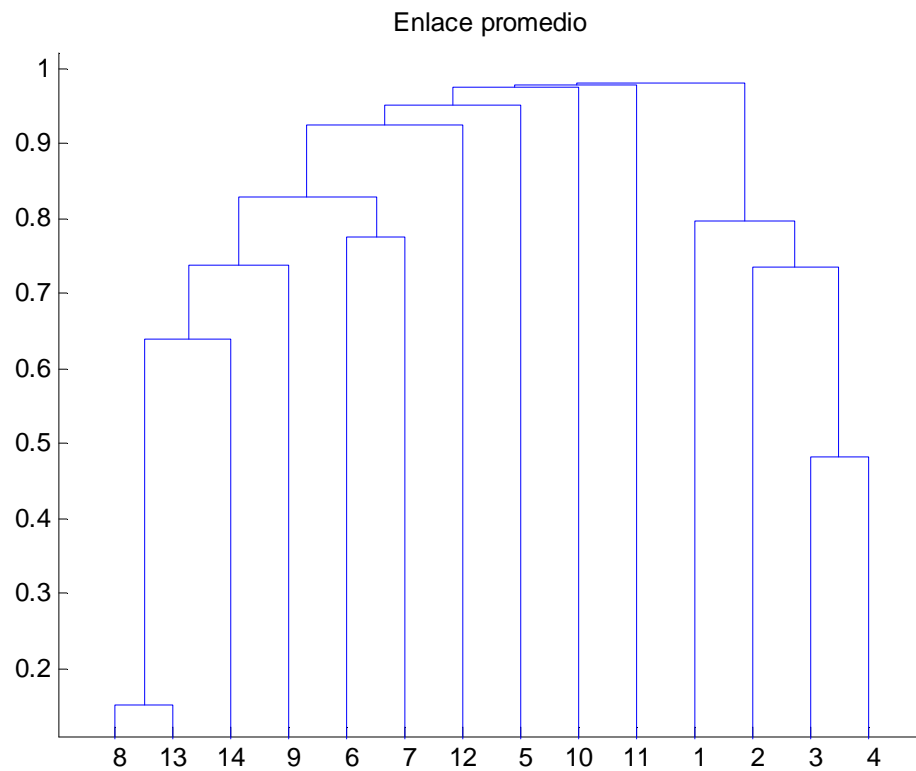
- DU2, 2007



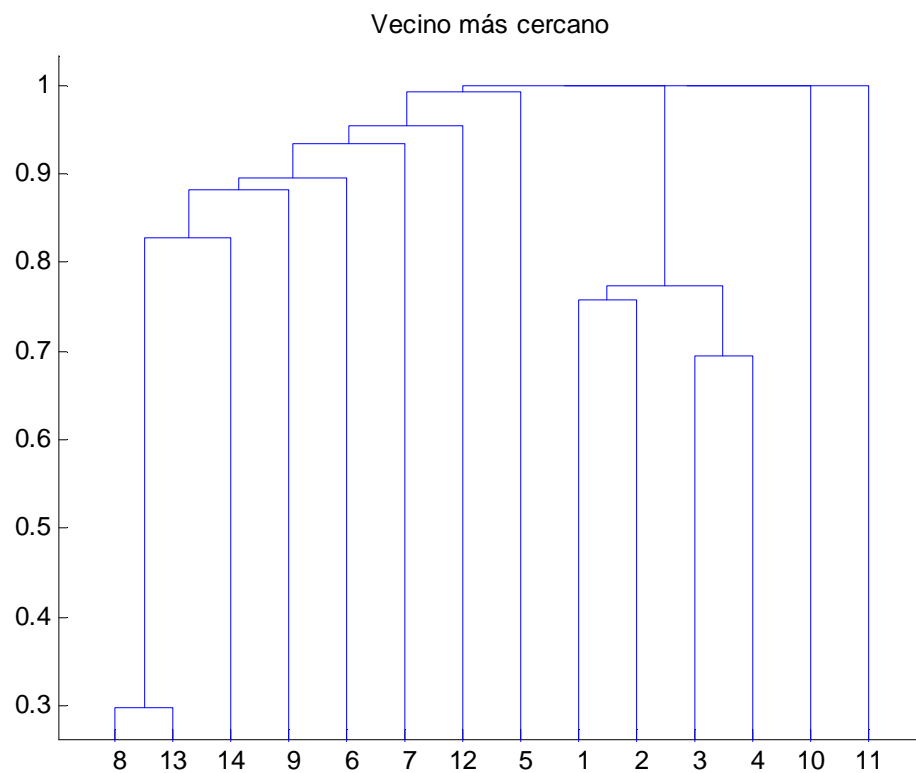


- Dcramer, 2008



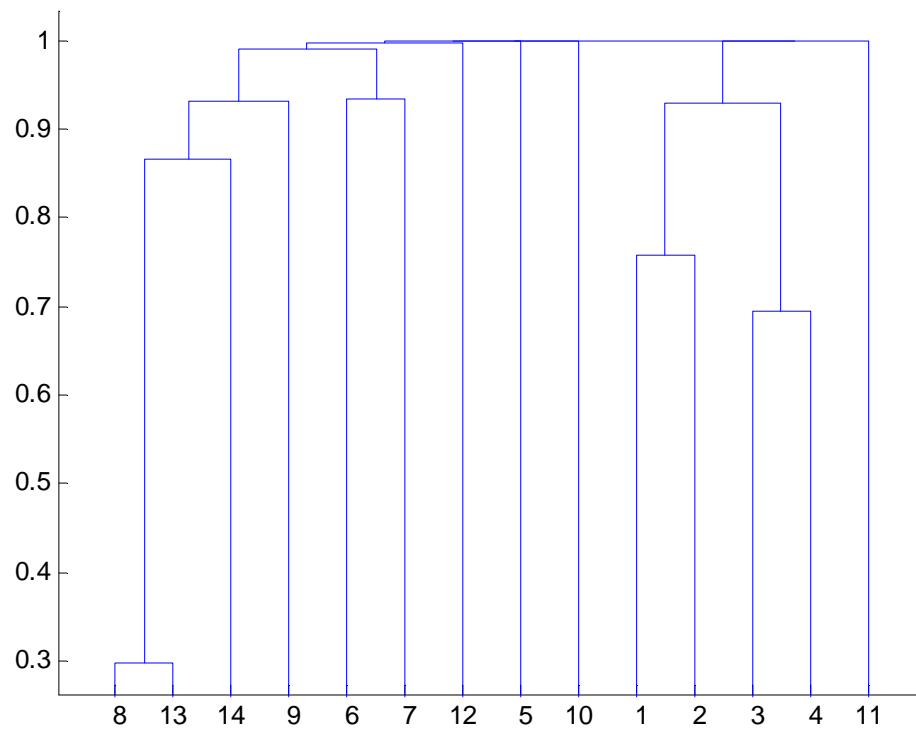


- DCI, 2008

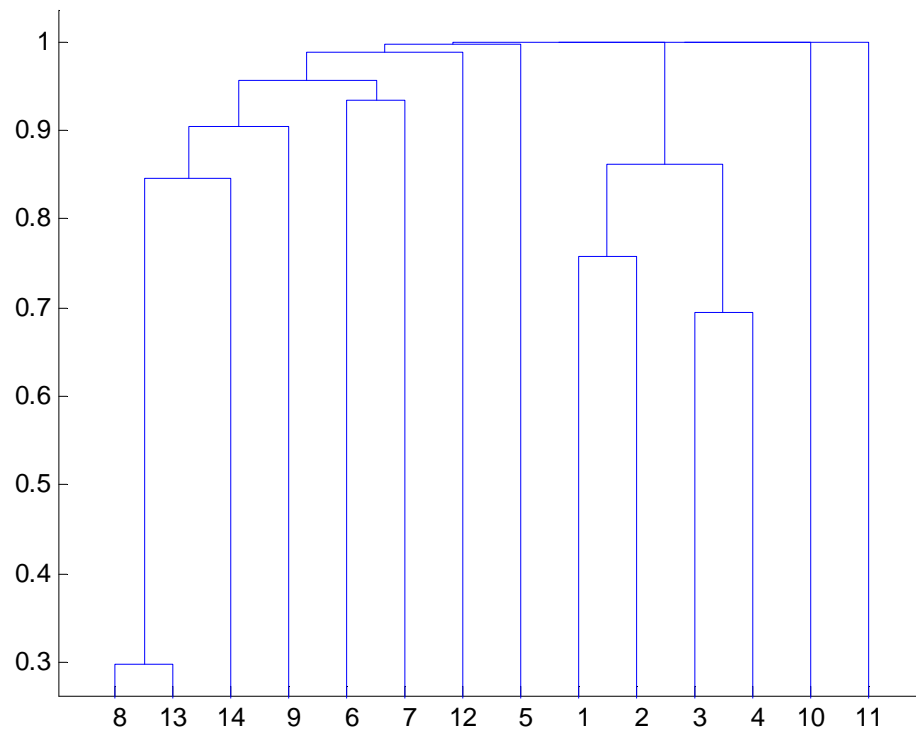




Vecino más lejano

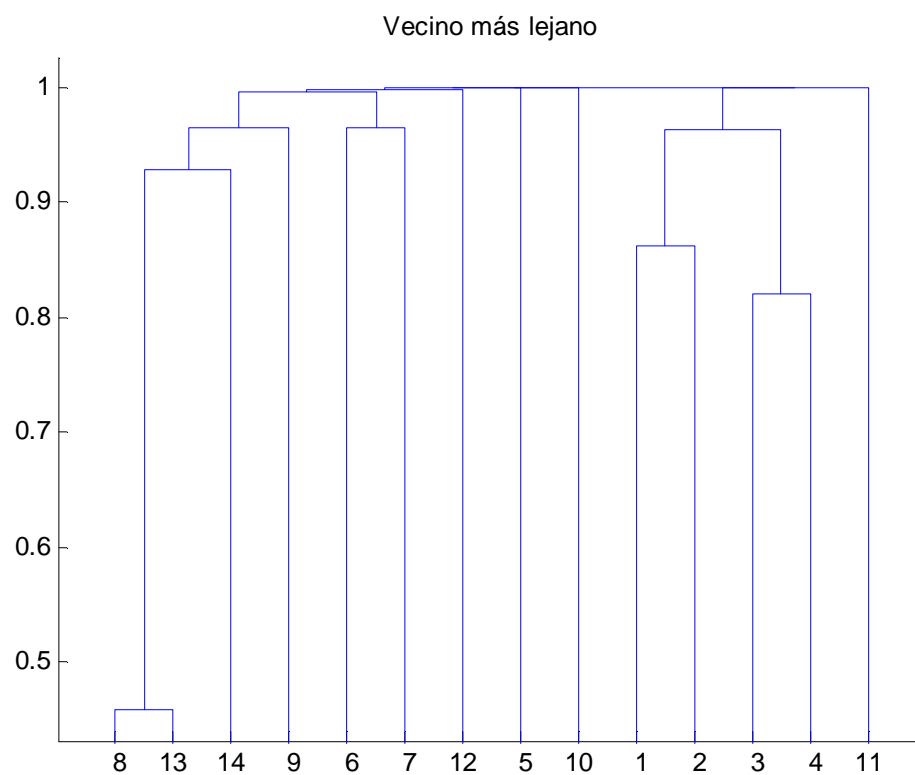
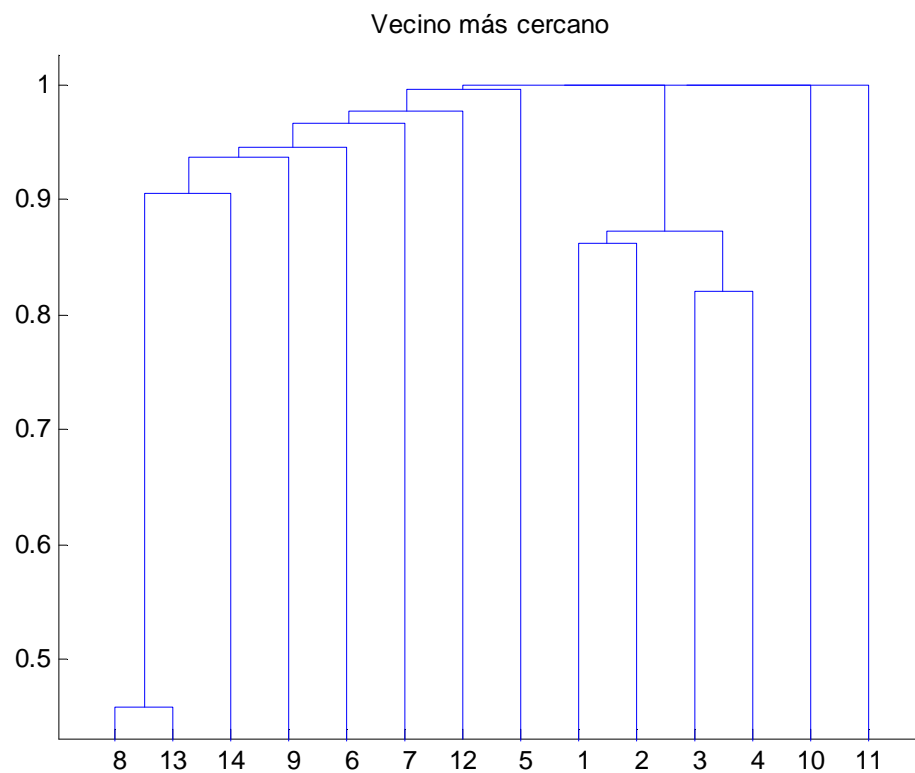


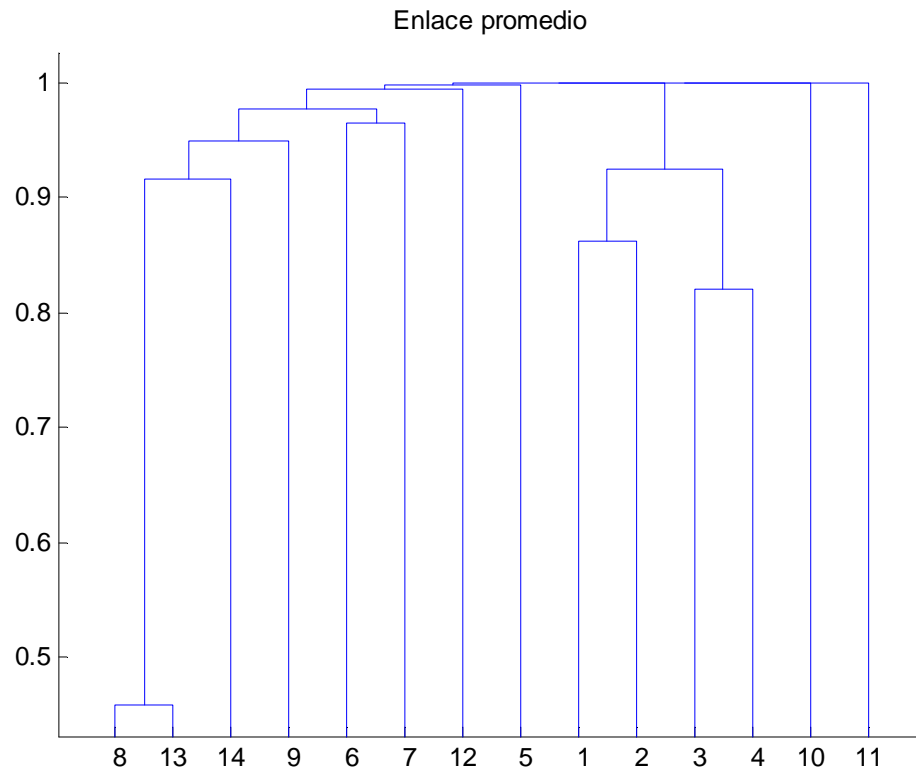
Enlace promedio



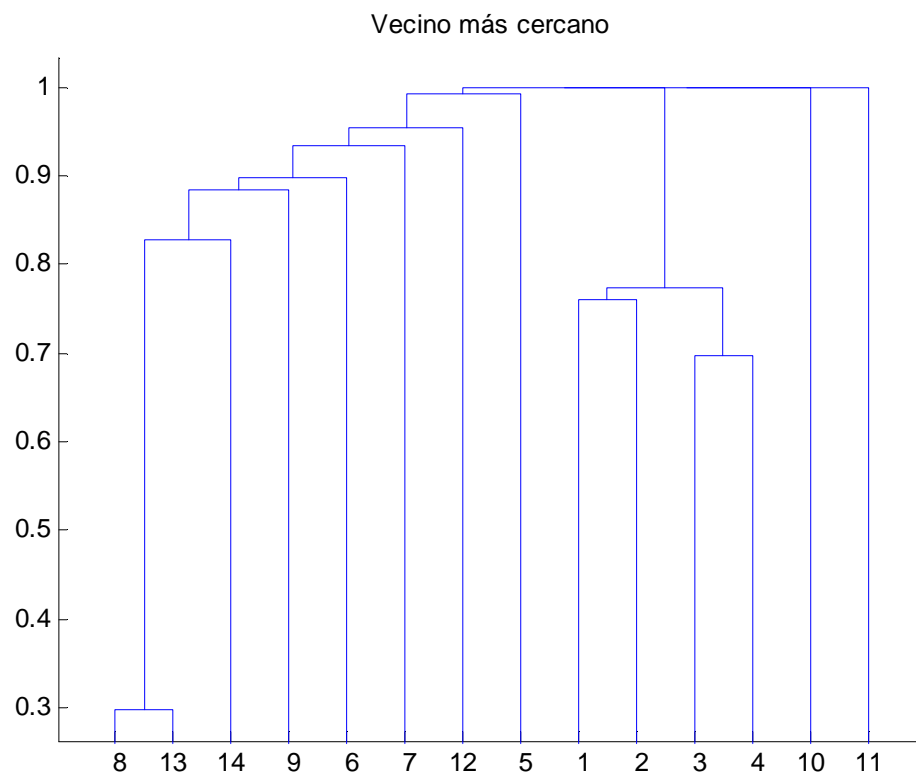


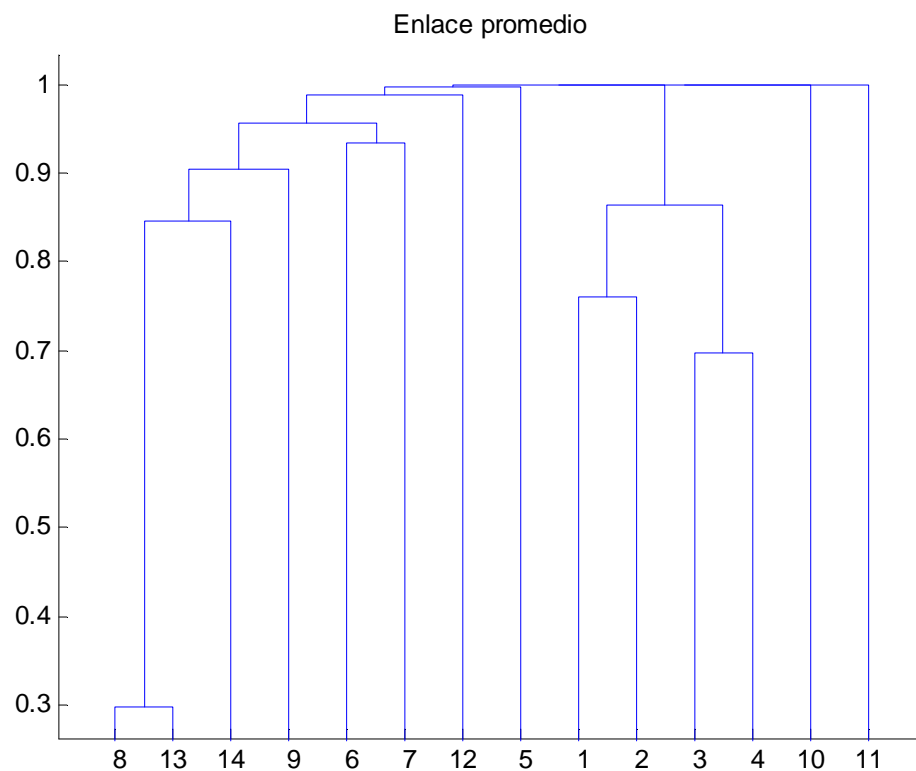
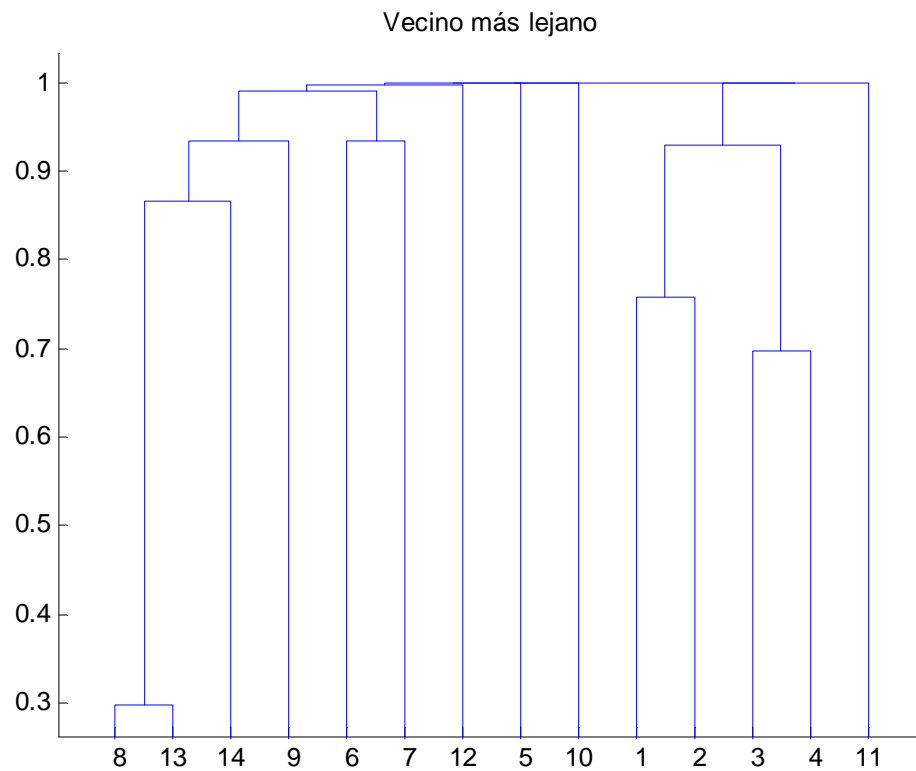
- DU1, 2008





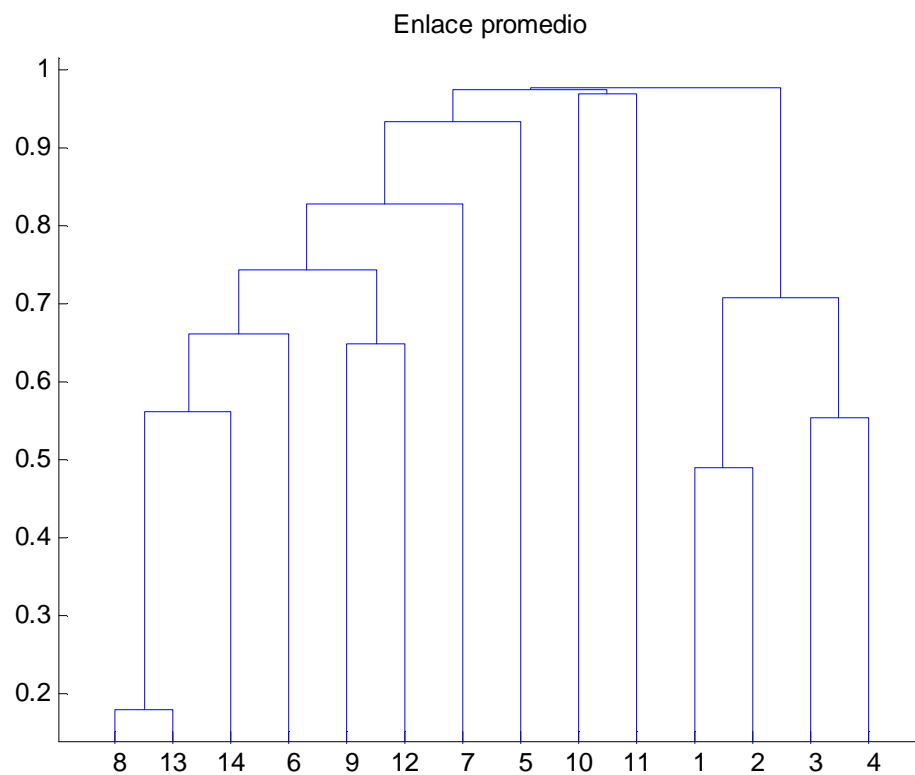
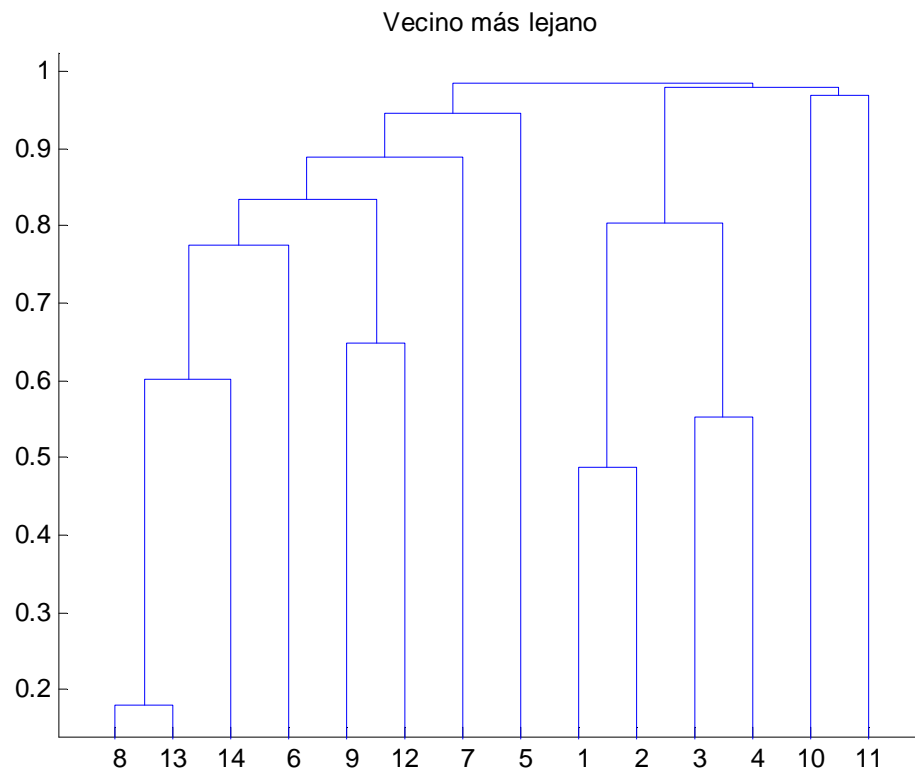
- DU2, 2008





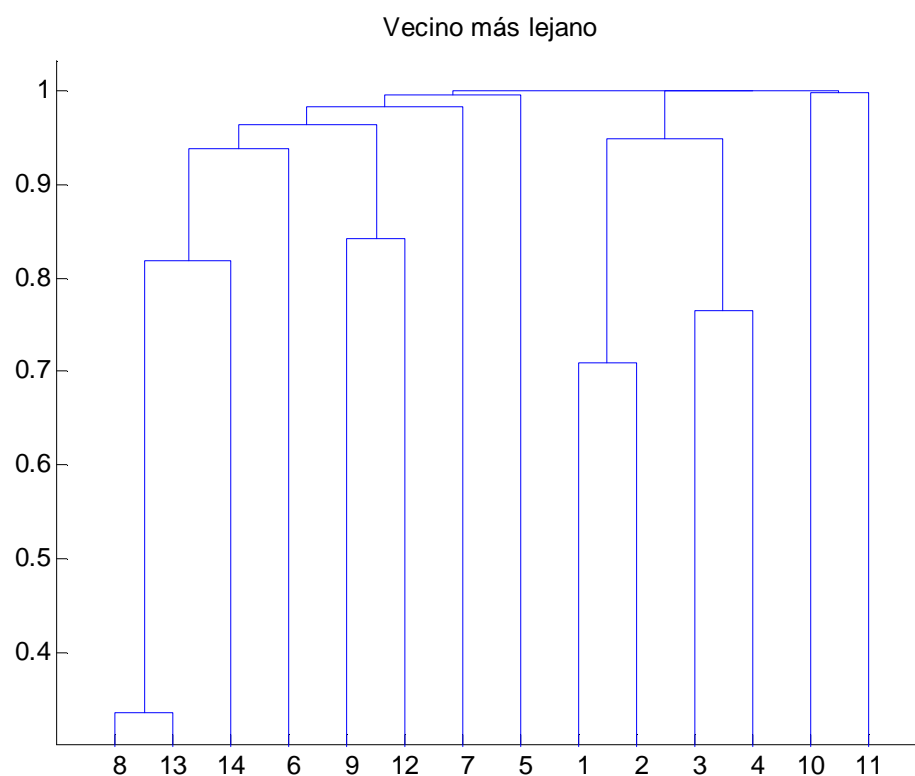
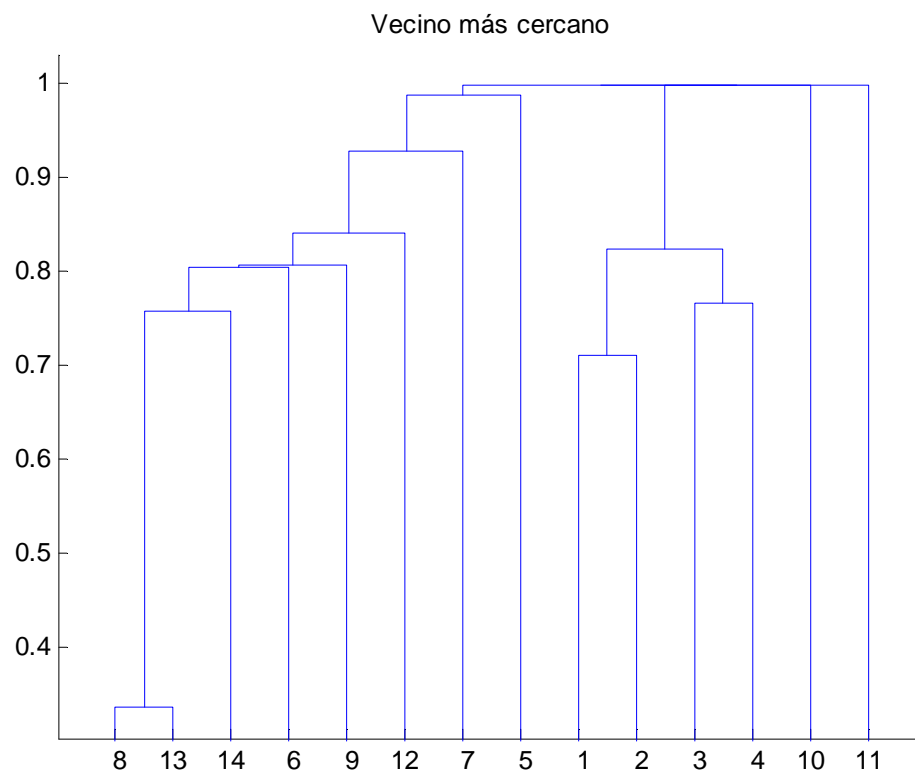


- Dcramer, 2009



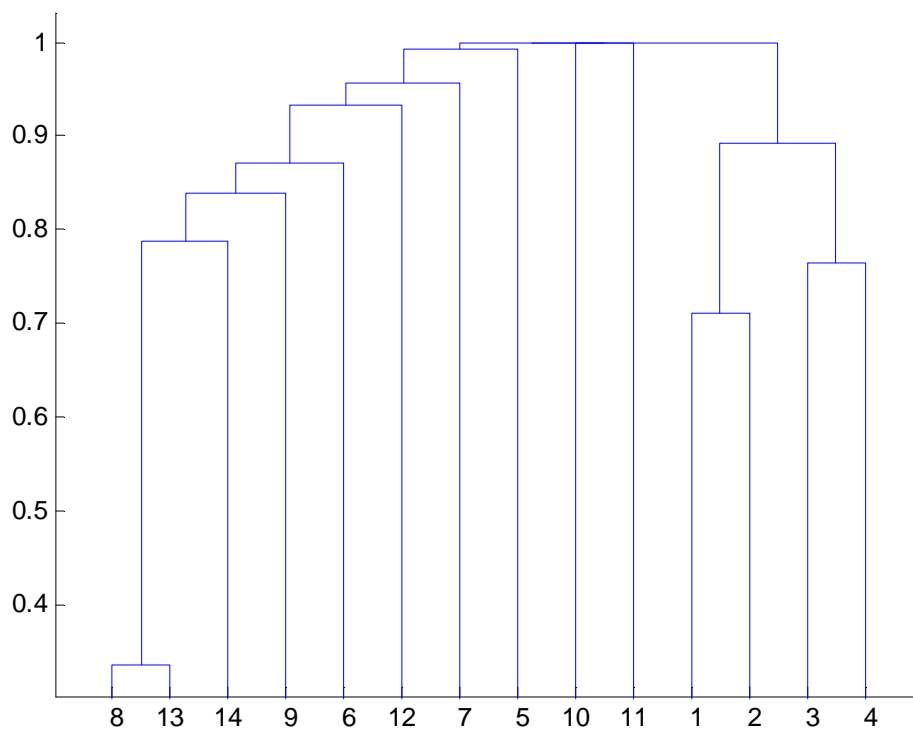


- DCI, 2009



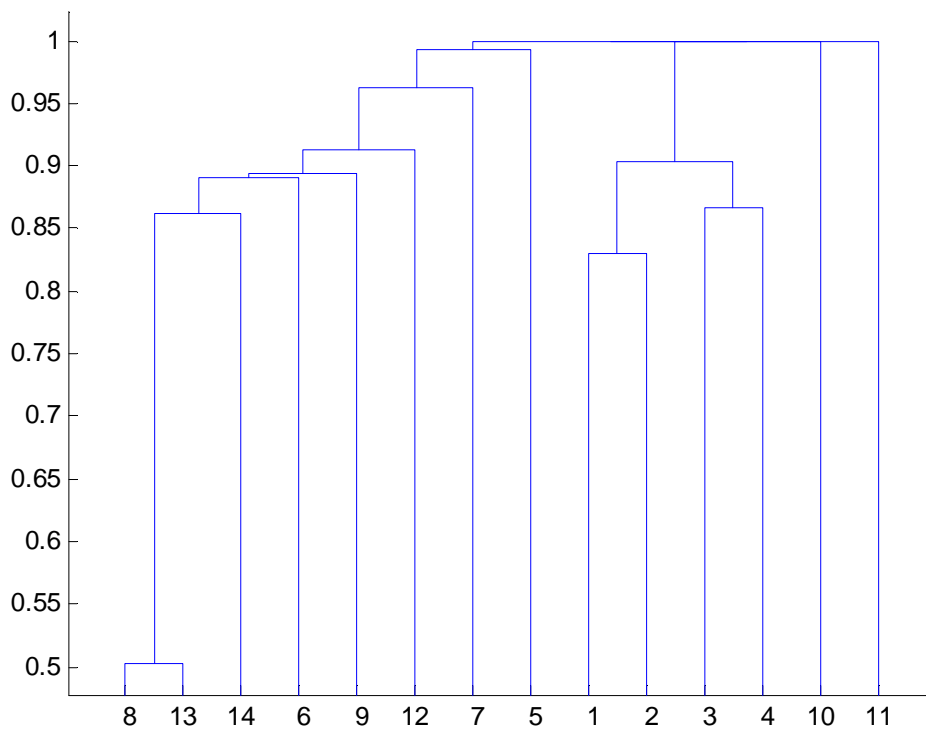


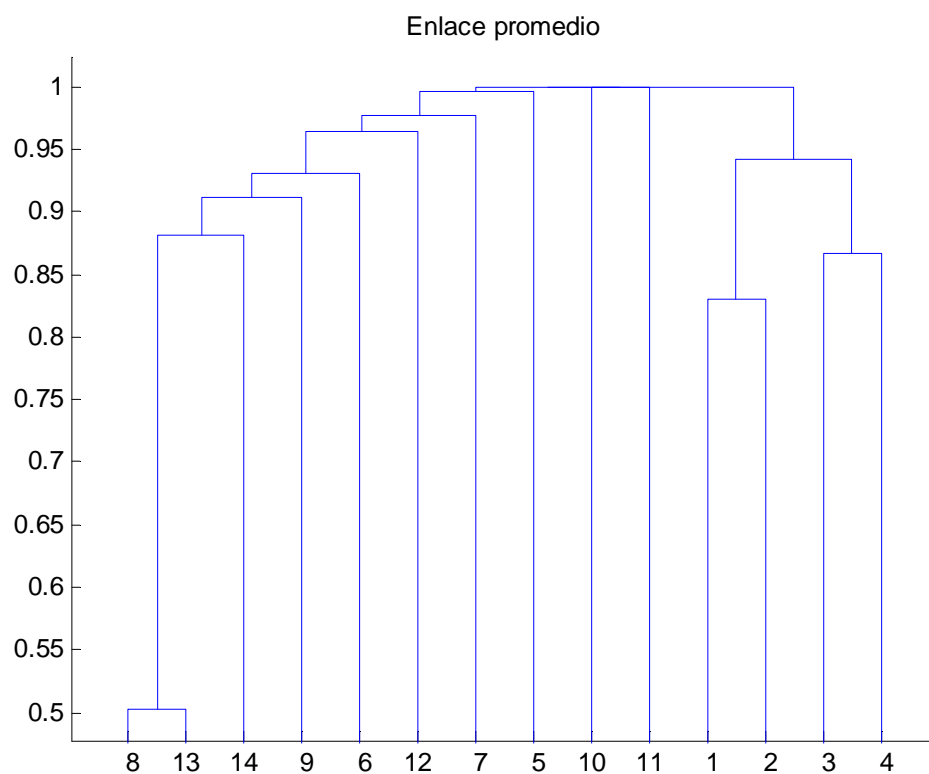
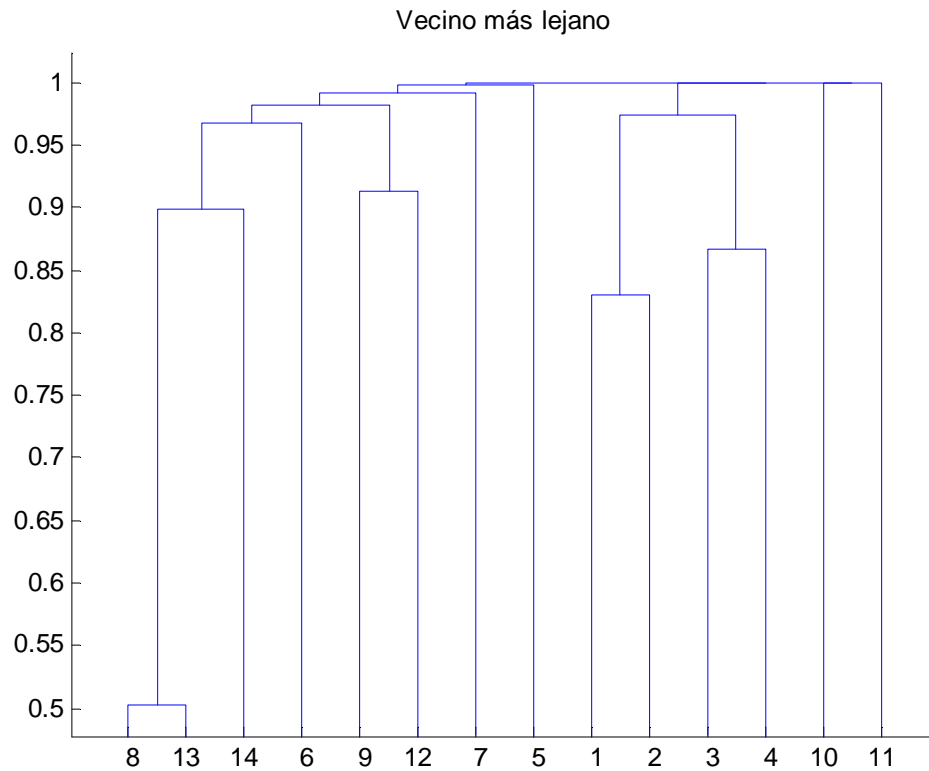
Enlace promedio



- DU1, 2009

Vecino más cercano

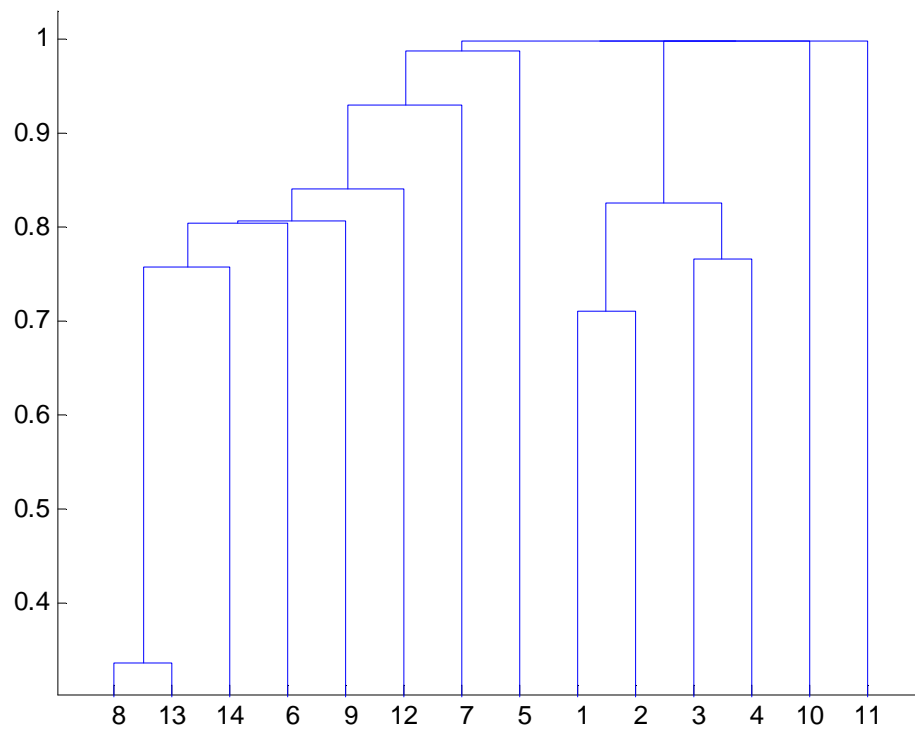




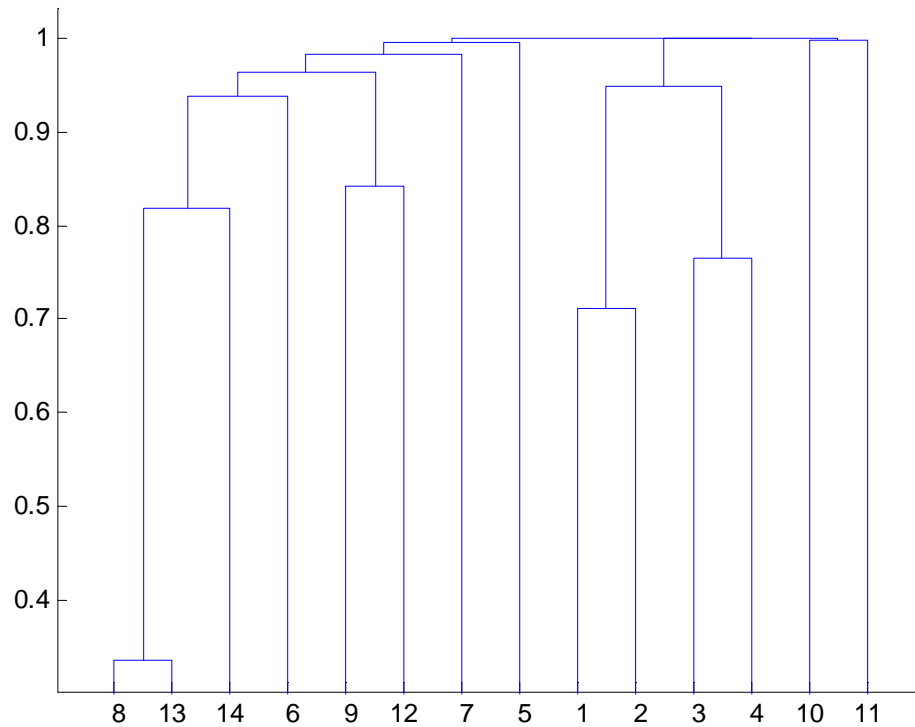
- DU2, 2009

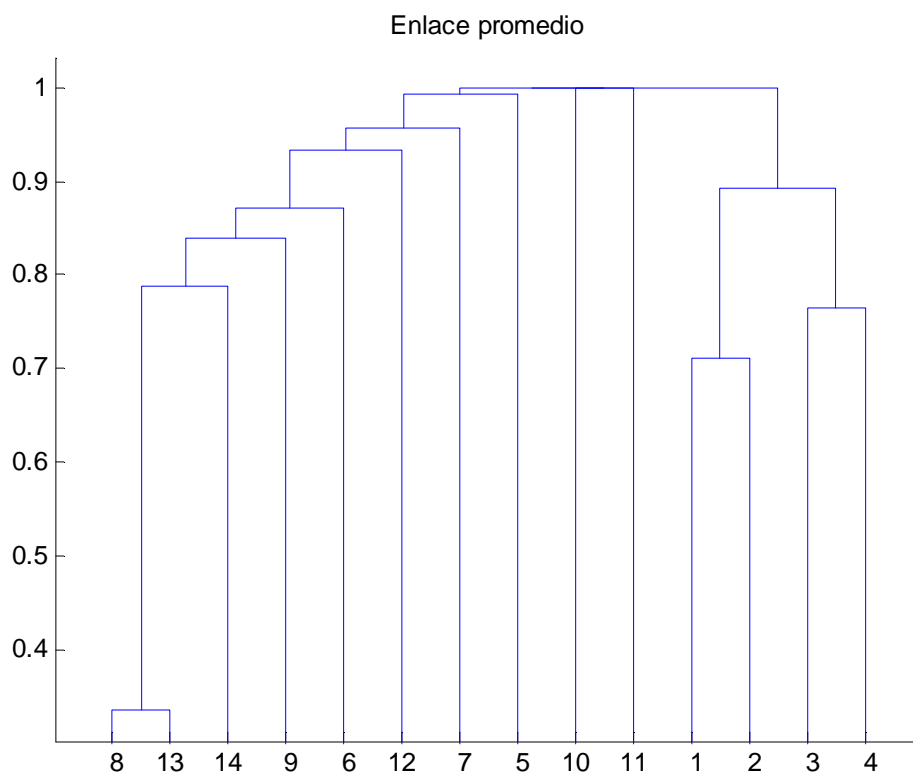


Vecino más cercano



Vecino más lejano





Cada mercado se corresponde con la siguiente numeración:

1. NSW
2. QLD
3. SA
4. VIC
5. AUSTRIA
6. DIN ESTE
7. DIN OESTE
8. FINLANDIA
9. NORDPOOL
10. OMEL
11. ONTARIO
12. OSLO
13. SUECIA
14. TRO

